

## ABSTRACT

Title of Dissertation: ASPECTS OF ONLINE REVIEWS AND THEIR EFFECTS IN CONSUMER DECISIONS

Jared Joseph Watson, Doctor of Philosophy in Business and Management, 2018

Dissertation directed by: Ralph J. Tyser Professor of Marketing, Amna Kirmani, Department of Marketing

Assistant Professor, Anastasiya Pocheptsova Ghosh, Department of Marketing

As online reviews increase in their prominence as a critical resource for consumers, investigation into their influences are necessary. This dissertation investigates various aspects of online reviews and how they affect consumers' decisions. In Essay I, the role average product ratings and review volumes are investigated, and the conditions by which the relative influence of these two summary attributes are defined. While average product ratings tend to be more diagnostic, and therefore more influential, than review volumes, when review volumes in a choice set are low or when the average product ratings are somewhat negative, consumers are more likely to rely on

review volumes to also inform their decisions. In Essay II, consumer responses to information veracity disclosures are investigated. Some websites now report when they have identified fake reviews, and we demonstrate that consumers may be overreacting to these alerts due to the salience of the information. Furthermore, we argue that consumers not only attempt to correct for fake reviews, but also punish brands who solicit them. Both essays have implications for theory and practice, while providing interesting avenues for future research.

ASPECTS OF ONLINE REVIEWS AND THEIR EFFECTS IN CONSUMER  
DECISIONS

by

Jared Joseph Watson

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park, in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2018

Advisory Committee:

Professor Amna Kirmani, Co-Chair

Assistant Professor Anastasiya Pocheptsova Ghosh, Co-Chair

Associate Professor Rosellina Ferraro

Associate Professor Michael Trusov

Professor Michael Dougherty

© Copyright by  
Jared Joseph Watson  
2018

## Table of Contents

CHAPTER I: INTRODUCTION	1
CHAPTER II: ESSAY I – Swayed by the Numbers: The Consequences of Displaying Product Review Attributes	4
CONCEPTUAL BACKGROUND	9
Attribute Diagnosticity	11
Bound versus Unbound Scales	13
Negative versus Positive Average Product Ratings	14
Low versus High Review Volumes	15
OVERVIEW OF STUDIES	17
STUDY 1	19
Design and Procedure	20
Results	20
Discussion	21
STUDY 2	22
Design and Procedure	23
Results	24
Discussion	26
STUDY 3	26
Design and Procedure	27
Results	28
Discussion	28
STUDY 4	29
Design and Procedure	30
Results	30
Discussion	32
STUDY 5	33
Design and Procedure	34
Results	35
Discussion	36
STUDY 6	37
Design and Procedure	38
Results	39
Discussion	40
STUDY 7	41
Design and Procedure	41
Results	42
Discussion	44
GENERAL DISCUSSION	45
Additional Studies	45
Managerial Implications	50
Conclusions and Future Directions	53
CHAPTER III: ESSAY II – Red Flag! The Consequences of Alerting Consumers to Fraudulent Reviews	56

CONCEPTUAL BACKGROUND	60
Review Processing	61
Persuasion Knowledge and Correction	61
Mediation via Perceived Average Rating Accuracy and the Desire for Justice	63
Review Reading Behaviors	65
Moderating the Effect of a Fake Review Alert	66
OVERVIEW OF STUDIES	66
STUDY 1	67
Data and Procedure	68
Results and Discussion	68
STUDY 2	70
Design and Procedure	71
Results	73
Discussion	77
STUDY 3	79
Design and Procedure	79
Results	80
Discussion	84
STUDY 4	84
Design and Procedure	85
Results	85
Discussion	90
GENERAL DISCUSSION	91
Theoretical Implications	92
Managerial Implications	92
Limitations and Future Research	93

#### List of Appendices

Appendix A (Essay I Stimuli)	95
Appendix B (Essay I Additional Studies)	105
Appendix C (Essay II Stimuli)	130

#### List of Tables

Table 1 (Essay I Summary of Designs and Measures)	170
Table 2 (Essay I Summary of Means)	171
Table 3 (Essay I Eye-tracking Fixation Transition Matrices)	172
Table 4 (Essay II Summary of Hypotheses Tested by Study)	173

#### List of Figures

Figure 1 (Essay I Conceptual Model)	174
Figure 2 (Essay I Example of Study Stimuli)	175
Figure 3 (Essay I Study 1 Preference Results)	176
Figure 4 (Essay I Study 3 Preference Results)	177

Figure 5 (Essay I Study 4 Preference Results)	178
Figure 6 (Essay I Study 4 Preference Results)	179
Figure 7 (Essay I Study 5 Preference Results)	180
Figure 8 (Essay I Managerial Implications Simulations)	181
Figure 9 (Essay II Study 1 Average Brand Ratings)	182
Figure 10 (Essay II Study 2 Brand Ratings)	183
Figure 11 (Essay II Study 2 Perception of Average Rating Accuracy)	184
Figure 12 (Essay II Study 2 Desire for Justice)	185
Figure 13 (Essay II Study 2 Average Valence of Reviews Read)	186
Figure 14 (Essay II Study 3 Brand Ratings)	187
Figure 15 (Essay II Study 3 Perception of Average Rating Accuracy)	188
Figure 16 (Essay II Study 3 Desire for Justice)	189
Figure 17 (Essay II Study 3 Average Valence of Reviews Read)	190
Figure 18 (Essay II Study 4 Brand Ratings)	191
Figure 19 (Essay II Study 4 Perception of Average Rating Accuracy)	192
Figure 20 (Essay II Study 4 Desire for Justice)	193
REFERENCES	194

## INTRODUCTION

This dissertation examines different aspects of online reviews and their effects in consumer decisions. Online reviews are proliferating at a tremendous rate, with most consumers now stating that online reviews are the most important product attribute in online purchase decisions (BrightLocal 2017). As such, it is important to understand how various aspects of reviews affect consumers' decisions, and outline the conditions by which some of these attributes may have conditional influences. To that end, we begin this dissertation by first investigating two numerical attributes of online reviews, average product ratings and review volumes. Furthermore, because online reviews are becoming such an influential tool, firms have begun to attempt exploiting consumers via fake reviews (Mayzlin, Dover, and Chevalier 2014; Luca and Zervas 2016). Thus, the second essay in this dissertation investigates how consumers respond when a website discloses that they have caught fake reviews being written for a specific brand.

In Essay I, we investigate how average product ratings and review volumes influence consumers' decisions when faced with a choice set in which there is no dominant option (i.e., when one option has a higher rating, but fewer reviews relative to another option). We argue that the diagnosticity (i.e., influence) of both average product ratings and review volumes are conditionally influenced by the other attribute, and as such, the choice between the higher-rated, fewer reviews option and lower-rated, more reviews option is dependent on the specific values of each attribute. While prior research has demonstrated the relative influence of both attributes, the findings are still debated (Floyd et al. 2014; You, Vadakkepatt, and Joshi 2015). By investigating the conditional effects of these attributes on choice, we help to rectify the divergent findings. We argue that average product ratings are inherently more diagnostic than review volumes due to



the bound versus unbound nature of their scales, respectively. Whereas average product ratings have stable scale boundaries (e.g., one to five stars), review volumes do not (e.g., zero to infinity). As such, review volumes are more susceptible to relative comparisons made within the choice set. We demonstrate how the relative diagnosticity of these attributes are a function of the review volumes contained within the choice set, and how this ultimately governs choice. We conclude Essay I with the theoretical implications as well as a series of simulations demonstrating the practical implications for managers.

In Essay II, we demonstrate the consequence of websites informing consumers that they have identified fake reviews for brands featured on their website. While a growing body of literature has investigated the characteristics of fake reviews (Mukherjee et al. 2013; Ott et al. 2013), as well as the firms which are likely to solicit them (Mayzlin, Dover, and Chevalier 2014; Luca and Zervas 2016), to the best of our knowledge, this is the first investigation into the effect of disclosing this information to consumers. While fake review alerts inform consumers that websites are monitoring the reviews for fraudulent information, we argue that the alerts also activate consumers' persuasion knowledge (Friestad and Wright 1994), leading to attempts to correct for perceived biased information, as well as justice against the brand when it is the source of the fake reviews. We demonstrate that fake reviews lead consumers to not only attempt correction in their perception of the brand, but also in the information that they acquire (i.e., the reviews they read). Furthermore, we show that reducing consumers' perceptions of inaccurate information attenuates their corrections. As such, this research holds relevance for website managers which provide reviews for their consumers.

In both essays, we demonstrate the consequences of review information in consumers' judgments and decisions. We argue that managers must carefully consider what information to provide consumers, and how to present it, in order to avoid biasing their consumers' decisions.

## CHAPTER II: ESSAY I

# Swayed by the Numbers: The Consequences of Displaying Product Review Attributes

Jared Watson, Dr. Anastasiya Pocheptsova Ghosh, Dr. Michael Trusov

With the rise of internet shopping, product reviews have gained prominence, with nearly 60% of consumers now saying that the average product rating is the most important product attribute in their purchase decisions. Within consumer reviews, 54% of consumers report paying attention to average product ratings and 46% to review volumes, suggesting a slightly higher weight being placed on the former attribute (BrightLocal 2017). Because average product ratings and review volumes play such a significant role in consumer behavior, marketing academics have tried to understand the processes by which consumers incorporate both pieces of information into their decisions (Godes and Mayzlin 2004; Chevalier and Mayzlin 2006; Clemons, Gao, and Hitt 2006; Liu 2006; Dellarocas, Zhang, and Awad 2007; Duan, Gu, and Whinston 2008; Chintagunta, Gopinath, and Venkataraman 2010; Zhu and Zhang 2010; Moe and Trusov 2011; Sun 2012; Ho-Dac, Carson, and Moore 2014). While the literature supports the positive influence of both average product ratings and review volumes on sales, their relative influence is still debated. Indeed, two recent meta-analyses arrived at opposing conclusions. Floyd et al. (2014) finds support for the claim that average product ratings are more influential than review volumes, while You, Vadakkepatt, and Joshi (2015) argues for the greater importance of review volumes.

While a few papers have investigated the interactive effects of various aspects of product reviews, they have stopped short of investigating the conditional effects of average product ratings and review volumes at various levels of each attribute. For example, Chintagunta, Gopinath, and Venkataraman (2010) examine the effect of a movie's average rating and review volume on its box office sales. While they account for the average ratings of competing movies, they do not include the review volumes for the

competing movies. Thus, they provide only a partial account of the effects of average ratings and review volumes in a competitive environment. Further, their analysis yielded no conditional effects of average ratings with a movie's own review volume. Rather than investigate reviews at the aggregate level, Chen, Dhanasobhon, and Smith (2008) investigated the influence of individual reviews on book sales. They found that reviews which received more "helpful" votes from other consumers were more predictive of sales relative to other reviews, and this was stronger for less, versus more, popular books. Lastly, Khare, Labrecque, and Asare (2011) investigated additional review attributes, and found that the ratings dispersion (i.e., the distribution of individual review ratings) differentially impacted sales of negatively- and positively-rated products, but only when review volumes were "high". Negatively- (positively-) rated goods benefited from wide (narrow) dispersions in consumer evaluations. Taken together, these results demonstrate conditional effects of both average product ratings and review volumes relative to other attributes, and thus, warrant investigation as to the conditional effects of these attributes relative to each other.

The goal of this paper is to address this gap in the literature and specify the conditions by which the interactive effect of average product ratings and review volume on consumer preference between options takes place. Theoretically, we draw from and contribute to several literatures. First, prior work in numerical cognition has investigated the influence of expanding versus contracting bound numeric scales of a single attribute on product evaluations (Bagchi and Li 2010; Pandelaere, Briers, and Lembregts 2011; Monga and Bagchi 2012; Schley, Lembregts, and Peters 2017), we extend this work by demonstrating the influence of multiple attributes that exist on both bound (e.g., average

product ratings: one to five stars) and unbound (e.g., review volume: zero to infinity) scales. Second, prior literature on attribute diagnosticity has demonstrated the context-dependent weight of attributes as a function of their perceived representativeness of product quality (Feldman and Lynch 1988; Herr, Kardes, and Kim 1991; Purohit and Srivastava 2001), we demonstrate that average product ratings are inherently more diagnostic than review volumes, though this can difference can be attenuated based on high levels of review volumes. In demonstrating this boundary condition, we can outline the conditions by which consumer reliance on both bound and unbound scales in their judgments varies.

Substantively, we also contribute to the domain of online product reviews by providing a comprehensive examination of the interactive effects of average product ratings and review volumes in a multi-option choice set. Whereas prior literature has addressed the influence of both average product rating and review volumes on evaluations of individual choice options (for a notable exception, see: Chintagunta et al. 2010), we examine the influence of both attributes in a competitive choice set. Increasingly, retailers are providing consumers with choice sets rather than individual choice options (e.g., product search pages, “recommended for you” lists, etc.), so consumers often encounter the average product ratings and review volumes of more than one option simultaneously. Thus, in our investigation, we examine the condition by which consumers prefer a higher-rated, fewer reviews choice option relative to a lower-rated, more reviews choice option. In doing so, we can isolate the relative diagnosticity of these two attributes, and the conditions by which this changes.

Consider the following scenario. Imagine searching for a new blender online. You see two comparable choice options that have the specifications you require. While one choice option has a higher rating but fewer reviews (e.g., 3.5 out of 5.0 based on 10 reviews), the other choice option has a lower rating but more reviews (e.g., 3.2 out of 5.0 based on 50 reviews). What is the relative diagnosticity of the average product ratings versus the review volumes as a signal of product quality? Would these diagnosticities, and ultimately your decision, change if the review volumes were instead 310 and 350, respectively? And what if the average product ratings were 4.5 and 4.2, respectively?

Managerially, these are important questions to study. Choice sets in which consumers face non-dominant options, in which one option is superior on the average product rating while another is superior on the review volume, are very common. Using secondary data of Amazon.com products, we compared the average ratings and review volumes of over 2.5 million products, across 24 product categories, to the choice options featured in the “customers who viewed this item also viewed” recommendation bar. This served as a proxy for the choice set of each product. Our analysis demonstrated that 79% of the products featured a higher rating, but fewer reviews than at least one option in their choice set (see Appendix B1 for additional details of our analysis). Thus, our investigation into the interactive effects of average product ratings and review volumes in a choice set holds not only theoretical relevance, but also practical relevance given how often consumers likely face this decision.

In the following sections, we develop our conceptual framework and the hypotheses to test it. We then test our hypotheses in a series of studies, before concluding

with the managerial and theoretical implications of our findings, and directions for future research.

## **CONCEPTUAL BACKGROUND**

### *The Diagnosticity of Attributes as Signals of Product Quality*

Consumers infer product quality from multiple product attributes when making choices (Slovic 1966; Slovic and Lichtenstein 1971; Rao and Monroe 1988, 1989; Richardson, Dick, and Jain 1994; Kirmani and Rao 2000). Slovic and Lichtenstein (1971) proposed the concept of attribute diagnosticity, and demonstrated how consumers differentially utilize multiple attributes in their decision as a function of the degree to which each attribute separates the available choice options relating to their perceived quality. Thus, attributes which have a greater degree of diagnosticity hold more influence in consumer decisions relative to those which are less able to separate the perceived quality options. The accessibility-diagnosticity framework built upon these findings, and demonstrated that the diagnosticity of attributes is dynamic in nature, leading to context-dependent rather than fixed levels of attribute diagnosticity (Feldman and Lynch 1988; Lynch, Marmorstein, and Weigold 1988). For example, Lynch, Marmorstein and Weigold (1988) demonstrated that when attributes were easily recalled, they were more diagnostic, and therefore indicative of choice, than attributes which were difficult to recall. Thus, it is not the inherent diagnosticity of an attribute that dictates choice, but rather the diagnosticity of each attribute at the moment of choice.



One context-dependent influence on diagnosticity is the valence of the available attributes. For example, Purohit and Srivastava (2001) demonstrated that manufacturer reputation is considered a primary diagnostic cue whereas product warranty is not. As a result, when a manufacturer has a very poor reputation, product warrant does not influence the evaluation of the product. However, if the manufacturer has an acceptable reputation, warranty is used as a secondary diagnostic attribute, where a longer warranty increases product evaluations. Thus, the value of one attribute can affect the diagnosticity of another attribute, leading to conditional effects of the attributes.

A related stream of literature, attribute evaluability, has also demonstrated context-dependent effects of attributes on preferences (Hsee 1996; Hsee et al. 1999; Hsee 2000; González-Vallejo and Moran 2001; Hsee and Zhang 2010). For example, Hsee (1996) asked participants to evaluate two dictionaries: one with 10,000 entries and in perfect condition, while the other had 20,000 entries with a torn cover. When evaluated independently, the former dictionary was preferred, but when evaluated simultaneously, preference for the latter dictionary increased. Said another way, independent evaluation led to greater diagnosticity of the book's cover whereas simultaneous evaluation increased the diagnosticity of the number of entries. Thus, product quality was judged more on the functional rather than aesthetic value of the good in the latter situation.

Taken together, these literatures support the claim that the diagnosticity of an attribute as a signal of product quality is not fixed. Consumers often evaluate products based on the values, and availability, of the attributes in the choice set, with little thought to attributes that are not accessible. Next, we apply these findings to our context: the usage of average product ratings and review volumes in choices between options.

### *Diagnosticity of Average Product Ratings and Review Volumes*

Most web sites provide consumers with more information than they can process. As such, consumers prioritize this information based on the diagnosticity of it in relation to their decision. Although, consumer reviews on most sites include the aggregate information (average product ratings and review volume) as well as the disaggregate information (individual ratings and text), the information, the aggregate and disaggregate information is often presented in two separate areas. Whereas aggregate information is presented on the search page, then again at the top of the individual product page, disaggregate information is often presented at the bottom of the individual product page or on another page entirely (i.e., accessed by a link on the product page). Thus, it is a more effortful process for consumers to access disaggregate information, and they are only likely to do so when highly motivated. For example, when choosing books to read, tastes are highly heterogeneous, and so aggregate information may not be diagnostic for a consumer. Rather, they may seek individual reviews to investigate fit between the reviewer and themselves (e.g., for a mystery book, a mystery fan may find the review from another mystery fan more diagnostic than someone who prefers biographies). Recent research on the topic has begun to investigate the influence of individual reviews on consumer decisions (Ludwig et al. 2013; Villaroel et al. 2017), and demonstrated that affective content contained in the individual reviews influence purchase conversion rates above and beyond the effects of the aggregate information, such as review volume (Ludwig et al. 2013). Given the separate structures of aggregate and disaggregate review information, as well as the individual differences associated with the diagnosticity each information source provides, investigating the relative influence of aggregate and

disaggregate review information would be challenging without compromising external validity (e.g., without asking participants which reviews they are reading and why, which is not something observed in industry).

Conversely, the aggregate information has convincingly demonstrate the sizeable influence of average product ratings and review volumes on consumers (Chevalier and Mayzlin 2006; Liu 2006; Duan, Gu, and Whinston 2008; Chintagunta, Gopinath, and Venkataraman 2010), yet as previously mentioned, their relative effects are still debated (Floyd et al. 2014; You, Vadakkepatt, and Joshi 2015). Complimenting this work, we argue that our investigation into average product ratings and review volumes can rectify some of the contrary findings by demonstrating the conditional influences whereby average product ratings and review volumes are more versus less influential.

In this paper, we limit the scope of our analyses to examining the relative diagnosticity of average product ratings and review volumes to increase the validity of our findings. From a theoretical perspective, accessing and interpreting disaggregate reviews is a more effortful process than accessing and interpreting aggregate reviews. As such, our focus is on rectifying findings from prior literature rather than integrating effort into the framework and providing additional findings which deviate from prior work. Second, consumers receive massive amounts of information online and are likely to engage in frugal processing of the information to minimize cognitive effort. As such, our paper begins with the least cognitively-demanding review attributes, to provide a foundation on which future research can provide greater examination of various levels of disaggregation (distributions, ratings trends, individual reviews, etc.).

We argue that average product ratings are inherently the most diagnostic due to their bound scale which allows consumers to quickly determine perceived product quality. Average ratings are generally presented on scales with two endpoints (e.g., 1 – 5 stars) so the inference process for any value on that scale is relatively easy. In comparison, review volumes are presented on unbound scales, where the minimum number of reviews is zero, but the maximum possible number of reviews is infinity. As such, the diagnosticity of review volumes is context-dependent based on what consumers perceive to be relevant review volumes. This leads to a lower inherent diagnosticity of review volumes as consumers may not have stable reference points, unlike with the average product ratings. However, in a choice set, consumers can compare the product review volumes to each other to shape their diagnosticity, leading to context-dependency as a function of the other review volumes. Consistent with our proposition, recent work by De Langhe, Fernbach, and Lichtenstein (2015) argues that the average product rating is the strongest indication of a products' objective quality. While this finding is not without debate (Kozinets 2016; Winer and Fader 2016), it does converge with the notion that consumers' expectations of average product rating as a diagnostic cue is warranted.

#### *Bound versus Unbound Scales*

Our argument for the relative diagnosticity of bound versus unbound scales is based on literature investigating numerical cognition. In general, numbers and calculations that are easier to process positively improve brand evaluations and product promotions (King and Janiszewski 2011). Applying this to our context, this finding would suggest that attributes on bound versus unbound scales would be more influential in decisions. Consistent with this view, Chandon and Ordabayeva (2017) examined

consumer estimates of increasing versus decreasing food quantities. Results demonstrated that consumers were more accurate when estimating decreasing amounts (bound by zero and the starting value) versus increasing amounts (bound by the starting value and infinity). Revisiting the work of Hsee (1996), we could also view the attributes of a dictionary as existing on bound and unbound scales. Individually, the book's condition is bound by completely destroyed and perfect, whereas the number of dictionary entries is bound by zero and infinity. Thus, the diagnosticity of dictionary entries is weak without a reference point. However, in the simultaneous evaluation context, consumers could use the number of dictionary entries for both options to inform the diagnosticity of that attribute for the dictionaries' quality.

#### *Negative versus Positive Average Product Ratings*

In this work, we argue that average product ratings are inherently more diagnostic than review volumes, and the influence of review volumes on preference between choice options is context-dependent based on the valence of the average product ratings.

Consumers are motivated to avoid the acquisition of bad products to minimize post-decisional regret (Tsiros and Mittal 2000; Zeelenberg and Pieters 2007). As such, consumers are likely to engage in a more elaborate assessment of choice options when to do so (Brockenholt et al. 1991).

Consistent with view, consumers are known to exhibit a negativity bias where they attend to, and elaborate more on, available information in the presence of negative information (Ito et al. 1998; Baumeister et al. 2001; Rozin and Royzman 2001). As such, we argue that the diagnosticity of review volume is likely to increase when consumers encounter average product ratings that contain some negative product information (e.g.,

low or neutral ratings). Conversely, if the average product ratings are considered rather positive (e.g., high ratings), then the individual average product ratings will be more influential than the review volumes in choice.

These predictions are based on the assumption that consumers see some merit in choosing the products. If consumers find all available options to be unsatisfactory (e.g., very low product ratings), they are unlikely to waste cognitive resources on making a decision, thereby showing no influence of either the average product ratings or the review volumes on their preference, instead relying on random choice.

-----

Insert figure 1 about here

-----

#### *Low versus High Review Volumes*

Thinking back to the opening example, how would consumer preferences between choice options be differentially affected if the choice set featured volumes of 10 (310) and 50 (350)? Although fewer versus more reviews would actually provide less information about the products, we argue that their diagnosticity would actually appear greater. Holding the absolute difference in the review volumes constant, the relative difference appears larger in the low versus high review volume conditions. Prior literature has demonstrated that consumers attend more to relative versus absolute differences (Thaler 1980; Tversky and Kahneman 1985). For example, participants indicated a greater willingness to drive to a farther-away store to save \$5 on a \$15

relative to a \$125 purchase. The authors argued that this occurred as a function of Prospect Theory, in which values exhibit a diminishing effect of utility as they grow (Kahneman and Tversky 1979; Tversky and Kahneman 1985). Although both conditions could receive \$5 savings at the further store, the utility from it appeared greater when it represented a larger discount (when the value of the item was lower). Thus, applying these insights into relative value sensitivities, we would expect that choice sets featuring low versus high review volumes would increase the diagnosticity of the review volume attribute.

A change in the perceived diagnosticity of review volumes (relative to average product ratings) has direct implications for consumer decisions. Given a choice set with non-dominant options (e.g., one option has a higher rating while another has more reviews), an increase in the relative diagnosticity of review volumes to average product ratings would weaken preference for the higher-rated good. As the diagnosticity of review volumes approaches that of average product ratings, a joint influence of both attributes emerges. Because the diagnosticity of review volumes increases when the review volumes of a choice set are low versus high, consumers are less likely to choose a higher-rated good, and more likely to choose a lower-rated (i.e., lower quality) good. However, the effect of low review volumes on preference can be attenuated when the choice set features average product ratings that are both considered good (i.e., high average ratings), as consumers will engage in a less effortful process in the absence of negative information, thereby being less likely to incorporate secondary cues like review volumes into their decision. Given a choice set in which consumers face a tradeoff

between an option with a better average product rating or one with a greater review volume, we formally propose:

**H1:** *Preference for higher-rated, fewer reviews choice option will be weaker when average product ratings are low or neutral (versus high), and when the level of review volumes is low (versus high).*

**H2:** *Preference between choice options is mediated by the difference in perceived diagnosticity of average product ratings and review volumes.*

## OVERVIEW OF STUDIES

We test our predictions in a series of seven lab studies. Study 1 demonstrates the systematic shift in preference between choice options as a function of review volume levels, providing initial support for H1. Study 2 tests the generalizability of this finding by demonstrating the effect with an expanded choice set. Study 3 then tests H1 in its entirety by examining the interaction of the ratings valence level and review volume level on preference between options. Studies 4 and 5 then demonstrate two boundary conditions by examining how the ratings size difference between options and the inclusion of a scale boundary attenuates the effect of a positive valence, respectively. Finally, Studies 6 and 7 test H2, in that the difference in perceived diagnosticity of average product ratings and review volumes mediates the effect of review volumes on preference, using self-reported weights of attributes and consumers' visual attention (captured by eye-tracking equipment) on the attributes.



-----  
Insert tables 1 & 2 about here  
-----

### ***Studies Paradigm***

In every study, participants were asked to imagine that they were considering the purchase of a new product and had narrowed their choice set to two comparable choice options (four choice options in Study 2). Participants then saw the choice options side by side, with information about the brand name, price, average product rating, and review volume for each choice option presented beneath the product images. In each choice set (except for Study 2), choice option A always had a higher average product rating with fewer reviews and choice option B had a lower average product rating with more reviews. Other product attributes were not significantly different.

Specific values of average product ratings and review volumes varied between studies and between the choice sets within the studies to extend the generalizability of our results (see Table 1). Review volumes were chosen by selecting values at the lower and upper limits of the perceived average review volumes based on a pre-test ( $N = 182$ ) in which we had participants classify various review volumes along a continuum from “1 = Far Fewer than Average, 7 = Far More than Average”.

After viewing each choice set, participants were asked to indicate their relative preference between choice options on a 7-point scale (1 = Strongly Prefer Option A, 7 = Strongly Prefer Option B” (except for Studies 2 and 6 in which we used a dichotomous choice dependent measure). This measure anchored preference for the higher-rated, fewer

reviews choice option at “1” and the lower-rated, more reviews choice option at “7”. As such, a higher number on this measure would indicate a weaker preference for the higher-rated, fewer reviews choice option. Figure 2 provides an example of the stimuli participants would view.

-----

Insert figure 2 about here

-----

### ***Study 1: Effect of Review Volumes on Preference between Options at Various Review Volume Levels***

The purpose of this study was to test the review volume levels argument of H1. We argue that preference for the higher-rated, fewer reviews options is weaker when review volumes are low relative to high. To test this, we examine four increasing review volume levels, while keeping average product ratings constant across conditions. We also include a fifth condition in which the review volumes are absent from the information provided to participants. This provides a condition in which preference is based largely on average product ratings, and comparing the other conditions to this one allows for an initial test of the diagnosticity of review volumes (H2). Finally, to demonstrate the robustness of this effect, we replicate it across five product categories where the brands, prices, average product ratings, and review volumes all slightly vary for each product to avoid any demand effects from specific values.

*Participants and design.* Two-hundred and fifty participants ( $M_{\text{age}} = 31.35$ ; 31% female; Amazon mTurk sample; \$0.50 payment) were randomly assigned to one of five review volume levels conditions: low (e.g., 8 vs. 64 reviews), moderate (e.g., 72 vs. 128 reviews), moderately high (e.g., 201 vs. 257 reviews), high (e.g. 456 vs. 512 reviews), or control (i.e., review volumes absent), in a between-subjects design. Within-subject, each participant viewed five product choice sets. The sample size was determined based on a 50-subject rule-of-thumb for online samples at the time the study was conducted.

*Choice sets.* For each of the five choice sets (headphones, microwaves, coffeemakers, speaker systems, and lounge chairs), participants would view two products which were nearly identical with the exception of their average product ratings and review volumes (see Table 1). In the control condition, review volumes were not displayed.

*Measure.* After viewing each choice set, participants were asked to “Please indicate your preference between options” on a 7-point scale (1 = Strongly Prefer Option A [higher-rated, fewer reviews option], 7 = Strongly Prefer Option B [lower-rated, more reviews option]). Thus, a lower score on this scale indicates greater preference for the higher-rated, fewer reviews choice option.

## *Results*

A 5 (review volume levels: low, moderate, moderately high, high, control) x 5 (product category: headphones, microwaves, coffeemakers, speaker systems, lounge chairs) repeated-measures ANOVA on preference yielded significant main effects of review volume level ( $F(4, 246) = 11.45$ ;  $p < .001$ ) and product category ( $F(1, 246) = 18.54$ ;  $p < .001$ ). The interaction was not significant ( $p > .10$ ). Because of this, we

collapsed across the product scenarios factor to simplify the reporting of results, though the same directional pattern held for all products. Supporting H1, planned contrasts demonstrated that preference for the higher-rated, fewer reviews choice option was significantly weaker in the low review volumes condition ( $M_{\text{low}} = 3.96$ ) compared to all other conditions ( $M_{\text{moderate}} = 2.93$ ;  $t(246) = -4.74$ ;  $p < .001$ ;  $M_{\text{m-high}} = 2.83$ ;  $t(246) = -5.12$ ;  $p < .001$ ;  $M_{\text{high}} = 2.63$ ;  $t(246) = -6.15$ ;  $p < .001$ ; and  $M_{\text{control}} = 2.81$ ;  $t(246) = -5.27$ ;  $p < .001$ ). Importantly, not displaying review volumes led to no significant difference in preferences relative to when review volumes were high ( $M_{\text{high}} = 2.63$  and  $M_{\text{control}} = 2.81$ ,  $p > .10$ ). This is consistent with our prediction that review volumes are less diagnostic, and therefore, less likely to influence preferences relative to average product ratings, when review volumes are high (H2).

-----

Insert figure 3 about here

-----

### *Discussion*

Study 1 demonstrated that consumers' preference for the higher-rated, fewer reviews option is weaker when review volumes in the choice set are low versus high. We replicated the effect in a follow-up consequential choice study, where participants were entered into a raffle to receive their preferred product option (a blender). Using the low and high review volume levels conditions from the main study, we observed the same shift in preference away from the higher-rated, fewer reviews option as the review

volume level decreased ( $M_{\text{low}} = 4.68$ ,  $M_{\text{high}} = 3.08$ ;  $F(1,104) = 19.10$ ;  $p < .001$ ), consistent with H1.

One interesting finding from this study demonstrated that preference for the higher-rated, fewer reviews option was not significantly different whether the review volumes were high or absent. This is consistent with the accessibility-diagnostics framework, as not displaying review volumes decreased the accessibility of review volumes, thereby decreasing their ability to influence consumer decisions. As such, this study provided initial evidence suggesting that managers may consider reducing the accessibility of low review volumes to aid consumers in choosing the higher-rated, and presumably higher quality, choice option.

### ***Study 2: Replication Using an Expanded Choice Set***

This study was designed to demonstrate that the effect of review volumes is robust with expanded choice sets. Prior research has demonstrated that large choice sets increase the use of noncompensatory decision strategies (Payne 1976; Johnson and Meyer 1984), such that consumers are more likely to choose choice options that are superior on one of the most important or easiest-to-differentiate attributes rather than incorporating multiple attributes. In the context of this research, we argue that when the review volume level is low, the diagnosticity of average product ratings and review volumes is relatively similar. As such, while consumers are more likely to use a noncompensatory strategy in the multiple choice option context, it is difficult for them to decide which attribute to rely solely on (ratings or volume), when review volume level is low (versus high). Thus, we expect a similar pattern of preferences to emerge as we observe in Study 1, where participants will be less likely to choose the highest-rated,

fewest reviews option when review volumes are low relative to high or absent, even in the context of the multiple options choice set where compromise options are available.

In addition, this study used a different dependent measure, choice deferral, to provide an initial test for H2. While many retailers attempt to provide consumers with multiple options to satisfy various consumer needs, an interesting consequence of this is that it increases the number of tradeoffs consumers must make with available attributes. Because of the increased diagnosticity of review volumes, when review volumes are low (relative to high or absent), the tradeoff between the review volumes and average product ratings attributes looms much larger. The need to make tradeoffs between choice attributes of similar importance increases choice difficulty (Chatterjee and Heath 1996; Dhar and Simonson 2003), which makes choice deferral more likely (Tversky and Shafir 1992; Dhar and Nowlis 1999; Etkin and Ghosh 2017). Thus, we expect the rate of choice deferral to be the highest in the low review volumes condition, relative to when review volumes are absent (i.e., a tradeoff is not salient) or when review volumes are high (i.e., the diagnosticity of review volumes is attenuated relative to average product ratings). Demonstrating this difference in deferral rates provides additional evidence for H2, which we directly test in Studies 6 and 7.

*Participants and design.* One-hundred and forty-four participants ( $M_{\text{age}} = 20.91$ ; 50% female; undergraduate sample; course credit) were randomly assigned to one of three review volume levels conditions (low, high, control) in a between-subjects design. The sample size was a convenience sample based on the undergraduate participants for a one-week time period.

*Choice set.* Participants viewed a choice set of four camping lamps, where options were nearly identical with the exception of their average product ratings and review volumes. While one choice option had the highest rating with the fewest reviews (e.g., 3.8, 5 reviews), another choice option had the lowest rating with the most reviews (e.g., 3.2, 61 reviews), and two other choice options in the middle were compromise choice options which were neither the highest, nor lowest on either attribute but were superior on one relative to the other compromise choice option (e.g., 3.4, 43 reviews and 3.6, 22 reviews) (see Table 1). Review volume levels were manipulated by either withholding the review volumes in the control condition or adding 300 reviews to the volumes reported above in the high review volume level condition.

*Measures.* To capture preference among the four choice options, participants were asked to “which option would you prefer?” and indicated their discrete choice of one of the four options. Thus, in this study our variable of interest was the choice of the highest-rated, fewest reviews choice option, rather than relative preference between options. To assess the likelihood of choice deferral, we asked participants “Are you more likely to purchase one of the available options or defer purchase, and look elsewhere?” and analyzed this as a binary measure. Lastly, to assess the need for more information, participants were asked “How would you classify the amount of information provided?” on a 7-point scale (1 = not enough information, 7 = too much information). A more difficult tradeoff would require more information to help participants make a decision, thus participants in the low review volume levels condition would be expected to require more information relative to those in the other conditions.

## *Results*

*Choice of the highest-rated, fewest reviews choice option.* A binary logistic regression, in which we dummy coded our review volume levels, yielded an omnibus effect of review volume levels ( $\chi^2(2) = 10.84; p = .004$ ). Consistent with H1, when review volumes were low, participants were significantly less likely to choose the highest-rated, fewest reviews choice option ( $P_{\text{low}} = 49\%$ ) relative to when review volumes were high ( $P_{\text{high}} = 78\%; \chi^2(1) = 7.09; p = .004$ ) or absent ( $P_{\text{control}} = 76\%; \chi^2(1) = 8.24; p = .008$ ). There was no significant difference in choice in the high and control conditions ( $p > .80$ ).

*Rate of choice deferral.* A binary logistic regression, in which we dummy coded review volume levels, yielded an omnibus effect of review volume levels ( $\chi^2(2) = 6.73; p = .035$ ). When review volumes were high ( $P_{\text{high}} = 53\%; \chi^2(1) = 4.30; p = .038$ ) or absent ( $P_{\text{control}} = 49\%; \chi^2(1) = 6.02; p = .014$ ) participants were significantly less likely to defer choice relative to when review volumes were low ( $P_{\text{low}} = 73\%$ ). There was no significant difference between high review volume and control conditions ( $p > .70$ ). A higher rate of choice deferral under low review volumes is consistent with prior work linking choice difficulty with choice deferral (Tversky and Shafir 1992; Dhar and Nowlis 1999; Etkin and Ghosh 2017). Consistent with our theorizing, when the level of review volumes was low (relative to high or absent), the diagnosticity of review volumes increased (H2), creating a more difficult choice involving the tradeoffs, ultimately increasing choice deferral.

*Need for additional information.* A one-way ANOVA of review volume levels on need for additional information yielded a marginal effect of review volumes ( $F(2,144) = 2.98; p = .054$ ). We define a p-value as marginally significant based on the rule-of-thumb



that it falls between .05 and .10. Planned contrasts further demonstrated that participants who encountered high review volumes felt they had significantly more information than those who had encountered low review volumes ( $M_{\text{high}} = 3.37$ ,  $M_{\text{low}} = 2.73$ ;  $t(144) = 2.43$ ;  $p = .016$ ), consistent with our predictions. Participants from whom review volumes were withheld were not significantly different from either of the other groups ( $M_{\text{control}} = 3.00$ ;  $p > .15$ ). While low review volumes increased the need for additional information consistent with our expectations, when review volumes were absent, participants felt no more need for additional information than when the review volumes were high. While not predicted, this result suggests that withholding review volumes from the list of attributes would not negatively impact consumers' perceptions of the amount of information they are provided with to make a choice.

### *Discussion*

Study 2 provided additional evidence for the effect of review volume by demonstrating that low review volumes shift preference away from the higher-rated choice options relative to when review volumes are high or absent, consistent with H1. Furthermore, this study provided evidence in support of H2, by demonstrating that choice deferral rate is the highest when review volumes are low relative to high or absent. Further, the need for additional information was greatest when review volumes were low, suggesting that consumers felt the need for more information when the tradeoff between average product ratings and review volumes was most salient.

### ***Study 3: Review Volume Effects at Various Valences of Average Product Ratings***

The objective of this study is to provide evidence for the entirety of H1. Using relatively neutral average product ratings (e.g., 3.0 – 3.8), our initial studies demonstrated

that consumers use both average product ratings and review volumes in their choices when review volume levels are low. H1 suggests that the effect of review volumes on consumer preferences can be attenuated when consumers make choices between positively-rated products. We argue that this occurs because positively-rated products can reduce consumers' desire to elaborate on more information when forming their decisions, thereby increasing their reliance on average product ratings relative to review volumes even further. Hence, we expect an attenuation of the effect of review volume levels when the choice set features positively-rated goods. To test this, in the next study, we manipulate the valence of average product ratings, while keeping the difference between ratings of options in the choice set the same.

*Participants and design.* Four-hundred and thirty-three participants ( $M_{\text{age}} = 20.28$ ; 46% female; undergraduate students; course credit) were randomly assigned to a condition in a 2 (review volume levels: low, high) x 3 (ratings valence levels: negative, neutral, positive) between-subjects design. The sample size was a convenience sample based on the undergraduate participants for a two-week time period.

*Choice set.* Participants saw a choice set of two blenders. Choice options were nearly identical with the exception of their average product ratings and review volumes (see Table 1). In the low (high) review volumes condition, participants chose between review volumes of 8 (408) and 64 (464), respectively. Ratings valence levels were manipulated by changing the first digit of the average product ratings for both choice options. Thus, the negative condition presented consumers with 2.x choice options, the neutral condition presented 3.x choice options, and the positive condition presented 4.x

choice options. After viewing the choice set, participants indicated relative preference on the same 7-point scale used in Study 1.

### *Results*

A 2 (review volume levels: low, high) x 3 (ratings valence levels: negative, neutral, positive) ANOVA on preference yielded main effects of review volume level ( $F(1, 427) = 17.26; p < .001$ ) and ratings valence level ( $F(2, 427) = 12.68; p < .001$ ), qualified by the predicted interaction ( $F(2, 427) = 3.58; p = .029$ ). Replicating prior studies, in the neutral valence level conditions, preference for the higher-rated, fewer reviews option was weaker when review volumes were low versus high ( $M_{\text{low}} = 4.38$ ,  $M_{\text{high}} = 3.38$ ;  $F(1, 427) = 14.28; p < .001$ ). As predicted in H1, a similar effect was present in the negative valence condition ( $M_{\text{low}} = 4.15$ ,  $M_{\text{high}} = 3.32$ ;  $F(1, 427) = 9.77; p = .002$ ). Furthermore, when the ratings valence level was positive, the effect of review volumes on preference was attenuated ( $M_{\text{low}} = 3.04$ ,  $M_{\text{high}} = 2.97$ ;  $F(1, 427) = .07; p > .75$ ). As expected, the effect of review volumes on preference is decreased, when the more diagnostic cue, average product ratings, is positive.

-----

Insert figure 4 about here

-----

### *Discussion*

This study provided full support for H1 by demonstrating that the effect of review volumes on preference between choice options is attenuated when the choice set features

only positive average product ratings. Literature on the negativity bias (Rozin and Royzman 2001) has demonstrated that consumers elaborate more on all available information in the presence of negative information. As such, when consumer encounter positive information (e.g., positive average product ratings), they elaborate less on additional information, thereby placing less weight on additional attributes in their decisions. Thus, consumers are likely to use average product ratings as their primary decision criteria relative to review volumes.

#### ***Study 4: Valence and Volume Effects across Small and Large Differences in Ratings***

The objective of this study was to examine another boundary of the valence effect on review volumes by examining how large difference in ratings between choice options attenuates the effect of a positive valence. In prior studies, we used relatively small differences between average product ratings (e.g., .2 – .4), in this study we expand the magnitude of difference to include .6 and .8. In doing so, we demonstrate a boundary condition of H1, by demonstrating that positive average product ratings attenuate the effect of review volumes only when the difference between average product ratings is small (versus large). By demonstrating this effect, we also provide additional evidence for the mediating role of attribute diagnosticity as large differences in average product ratings make the tradeoff amongst attributes more salient, leading consumers to rely on additional attributes to influence their decision.

To test these predictions, we created choice sets with four magnitudes of differences between choice options: .2 (e.g., 3.8 versus 3.6), .4 (e.g., 3.8 versus 3.4), .6 (e.g., 3.8 versus 3.2), and .8 (e.g., 3.8 versus 3.0). We did so across two different valences of product ratings: neutral and high (e.g., 3.x or 4.x). This resulted in a necessarily

unbalanced design because there are more differences of .2 within a level than .8, however, since there was no significant difference in effect within a given distance (for example, a difference of 4.2 versus 4.4 led to similar preference as a choice set that had 4.4 versus 4.6), we collapsed across those cells for the analysis. Thus, we collapsed across all small difference (.2-.4) and large difference conditions (.6-.8) to generalize our findings across relatively small and large difference in average product ratings.

*Participants and design.* Seven-hundred and five participants ( $M_{\text{age}} = 35.61$ ; 47% female; Amazon mTurk; \$0.50 payment) were randomly assigned to a condition in a 2 (review volume levels: low, high) x 2 (ratings valence levels: neutral, positive) x 2 (magnitude of ratings difference: small, large) between-subjects design. The sample size was determined based on balancing a 100-subject per cell rule-of-thumb for online samples at the time the study was conducted with cost considerations. This resulted in an a priori goal of 90 subjects per cell.

*Choice Set.* Participants saw a choice set of two headphones. Choice options were nearly identical with the exception of their average product ratings and review volumes as described earlier (see Table 1), and relative preference was measured on the same 7-point scale as in Studies 1 and 3.

## *Results*

A 2 (review volume levels: low, high) x 2 (ratings valence levels: neutral, positive) x 2 (ratings difference size: small, large) ANOVA on preference yielded significant main effects of review volume levels ( $F(1, 697) = 40.66$ ;  $p < .001$ ) and ratings valence levels ( $F(1, 697) = 6.70$ ;  $p = .01$ ), qualified by the interaction of review volume levels and difference size ( $F(1, 697) = 4.12$ ;  $p = .043$ ), and a marginal interaction of

review volume levels and ratings valence levels ( $F(1, 697) = 3.47; p = .063$ ). The main effect of review volume levels demonstrated that preference for the higher-rated, fewer reviews option was greater when review volumes were high ( $M_{\text{low}} = 3.63, M_{\text{high}} = 2.82$ ), consistent with prior studies. The main effect of valence demonstrated that preference for the higher-rated, fewer reviews option was greater when valence was positive relative to neutral ( $M_{\text{neutral}} = 3.37, M_{\text{positive}} = 3.08$ ), consistent with Study 3.

Consistent with our expectations, the review volume levels by ratings difference interaction indicated that the effect of the magnitude of the average product ratings difference on consumer preferences depended on review volume levels. When review volumes were low, consumers were less sensitive to the magnitude of differences in ratings between options ( $M_{\text{small}} = 3.62, M_{\text{large}} = 3.68; p > .85$ ), suggesting that any ratings difference was considered a tradeoff with the review volumes, consistent with H2 that proposes an increase in relative diagnosticity of review volumes when the level of review volumes is low. By contrast, when review volumes were high, consumers were more sensitive to the magnitude of differences in ratings between options, increasing preference for the higher-rated choice option, when the magnitude of difference between options was large versus small ( $M_{\text{large}} = 2.45, M_{\text{small}} = 2.98; p = .01$ ). This finding is consistent with H2, which proposes that the relative diagnosticity of average product ratings is greater when review volumes are high (versus low), thus increasing sensitivity to any difference in average product ratings.

The ratings valence levels by review volume levels interaction was consistent with the findings of Study 3. While low review volumes decreased preference for the higher-rated, fewer reviews choice option, the magnitude of this effect was larger in the

neutral relative to positive valence condition (neutral valence:  $M_{\text{low}} = 3.98$ ,  $M_{\text{high}} = 2.77$ ;  $p < .001$ ; positive valence:  $M_{\text{low}} = 3.32$ ,  $M_{\text{high}} = 2.66$ ;  $p = .012$ ). Note that the positive valence attenuated, rather than completely eliminated, the effect of review volumes on preference (as seen in Study 3). While this nominal difference could be a result of the specific stimuli used for each study, both studies demonstrated an attenuation of the influence of review volumes, consistent with our prediction, albeit to varying strengths.

-----

Insert figures 5 & 6 about here

-----

### *Discussion*

This study replicated the finding of Study 3 by demonstrating that positive valences attenuate the influence of review volumes on consumer decisions. Importantly, it also tested how a magnitude of difference in ratings between choice options changes consumer preference at different review volume levels. Our results suggested that average product ratings are more diagnostic than review volumes when review volume levels are high or when the average product ratings are positive, consistent with prior studies. Yet, this study also demonstrated that large ratings differences can attenuate the influence of positive ratings, leading consumers to utilize both average product ratings and review volumes to inform their decisions when there is a large difference in the average product ratings, even if they are both positive. Thus, a large ratings difference

can decrease the diagnosticity of positive average product ratings, demonstrating further importance for understanding the effect of review volumes on consumer decisions.

#### ***Study 5: Review Volume Effects at Ratings Scale Boundaries***

The objective of this study was to demonstrate a potential boundary effect of the influence of valence on review volumes. As we argued earlier in the paper rating scales have clear defined boundaries (e.g., 1.0 – 5.0) allowing for easier comparisons on that scales as compared to unbound scales, leading to increased diagnosticity of this attribute. In the next study, we propose and test the idea that endpoints specifically tend to be particularly diagnostic for consumer judgments.

We build this proposition on the work by Isaac and Schindler (2014) which demonstrates that consumers often form mental boundaries of ranked lists around the numbers that end in zeroes (e.g., “top 10” or “top 100”). Even if a list has more than 10 options, consumers will evaluate those within and outside of the top 10 differently. Since grouping of options affects types of comparisons that people make and final choices (Brenner, Rottenstreich, and Wood 1999), 10th option is perceived significantly more differently than adjacent options. Isaac and Schindler (2014) demonstrate this effect in the context of student rankings. Imagine students in a classroom ranked on performance. The 11<sup>th</sup>-ranked student is perceived to be significantly worse than the 10<sup>th</sup>-ranked student, however there is no difference in evaluation of the 11<sup>th</sup>- and 12<sup>th</sup>-ranked students. The “top 10” effect could be explained within the confines of Prospect Theory (Tversky and Kahneman 1992), where losses are shown to loom larger (i.e., be more important) than gains. Thus, if consumers’ point of reference is the top 10, 10<sup>th</sup> place would meet their standard while 11<sup>th</sup> place would be considered a significant loss. However, if they



are evaluating 11<sup>th</sup> and 12<sup>th</sup> place, both fall below their reference point of the top 10, diminishing sensitivity to the loss on this attribute and leading to an attenuation of the importance of their rankings.

Similarly, in the context of our research, we expect that when the rating of one product in a choice set lies at the boundary (i.e., has a rating of 1.0 or 5.0), the diagnosticity of average product ratings increases, attenuating the influence of review volumes. For example, in a choice set featuring ratings of 5.0 versus 4.7, consumers will be more influenced by the ratings rather than the review volumes, relative to when they view a choice set featuring 4.9 versus 4.6. While one could expect the same effect to occur at the negative boundary of the scale, the effect is likely to be smaller, consistent with the Prospect Theory, where a loss relative to the reference point of 5.0 is more significant than an equivalent gain relative to the reference point of 1.0.

*Participants and design.* Four-hundred and ten participants (51% female;  $M_{\text{age}} = 37.88$ ; Amazon mTurk sample; \$0.50 payment) were randomly assigned to a condition in a 2 (review volume level: low, high) x 2 (ratings valence level: negative, positive) x 2 (scale boundary included: no, yes) between-subjects design. The sample size was determined based on a 50-subject rule-of-thumb for online samples at the time the study was conducted.

*Choice set.* Participants saw a choice set of two hand mixers. Choice options were nearly identical with the exception of their average product ratings and review volumes (see Table 1). Valence and scale boundary were manipulated by the specific values used for average product ratings. The negative valence conditions were 1.3 versus 1.0 (boundary included) and 1.4 versus 1.1 (boundary not included). The positive valence

conditions were 5.0 versus 4.7 (boundary included) and 4.9 versus 4.6 (boundary not included).

*Measures.* After viewing the choice options, participants indicated relative preference on the same 7-point scale used in previous studies.

## *Results*

*Relative preference.* A 2 (review volume level) x 2 (average product ratings valence) x 2 (scale boundary included) ANOVA on relative preference between options yielded a significant main effect of valence ( $F(1, 404) = 3.98; p = .047$ ), a marginal main effect of the review volume level ( $F(1, 404) = 3.12; p = .078$ ), and a significant interaction between the review volume level and scale boundary inclusion ( $F(1, 404) = 4.47; p = .035$ ), qualified by the three-way interaction ( $F(1, 404) = 12.98; p < .001$ ). To explain the relationship between these factors, we will examine the two-way interactions between review volume levels and scale boundary at the negative and positive ends of the rating scale.

At the positive end of the scale (4.6 – 5.0), a 2 (review volume level) x 2 (scale boundary inclusion) ANOVA on preference yielded a marginal effect of the review volume level ( $F(1, 202) = 2.84; p = .093$ ) qualified by a significant interaction of the review volume level and scale boundary inclusion ( $F(1, 202) = 13.98; p < .001$ ). When the scale extreme was not included (i.e., 4.9 versus 4.6), low review volume level weakened preference for the higher-rated option with fewer reviews ( $M_{\text{low}} = 4.08, M_{\text{high}} = 2.71; F(1, 202) = 14.86; p < .001$ ), consistent with the results of our prior studies. By contrast, when the scale extreme was included (i.e., 5.0 versus 4.7), the effect of review volumes was attenuated ( $M_{\text{low}} = 3.48, M_{\text{high}} = 2.96; F(1, 202) = 2.09; p = .15$ ). Consistent

with Isaac and Schindler (2014) findings, this suggests that when products have perfect ratings (5.0), the ratings become significantly more influential in the decision process relative to when product have near perfect ratings (4.9).

At the negative end of the scale (1.0 – 1.4), a 2 (review volume level) x 2 (scale boundary inclusion) ANOVA on preference yielded no significant effects ( $p > .20$ ).

Across all negative conditions, preference averaged 3.64, where 4 indicate no preference between options. This effect was not predicted a priori and we discuss it next.

-----

Insert figure 7 about here

-----

### *Discussion*

This study demonstrated that the effect of review volume levels is attenuated when the scale boundary is included in the ratings for an option in the choice set. An interesting asymmetric valence effect emerged, such that at the negative end of the scale, review volume levels were attenuated regardless of whether or not the scale boundary was included. This could be a function of consumers making a choice between two unattractive options (Dhar and Sherman 1996), which is known to increase perceptions of choice difficulty. This finding is also consistent with the work of Purohit and Srivastava (2001), which demonstrated that when the primary diagnostic attribute is unattractive, subsequent attributes are not influential. To further explore this point, in a follow-up study we compared the effect of review volumes across very negative ratings (1.3 versus

1.0) and somewhat negative ratings (2.4 versus 2.1). Replicating the findings in this study, we found no effect of review volumes in the very negative condition ( $M_{\text{low}} = 3.51$ ,  $M_{\text{high}} = 3.14$ ;  $F(1, 83) = 1.13$ ;  $p = .29$ ). However, consistent with our findings of Study 3, we replicated the effect of review volumes in the somewhat negative condition ( $M_{\text{low}} = 3.89$ ,  $M_{\text{high}} = 3.16$ ;  $F(1, 85) = 3.75$ ;  $p = .056$ ). Further, consistent with the view that choosing between two extremely unattractive choice options increases choice difficulty, and potentially leads to the random choice between low valence options in the main study, rate of choice deferral was significantly higher in the very negative condition (45.9%) as compared to the somewhat negative condition (21.8%,  $\chi^2(1) = 10.93$ ;  $p = .001$ ). This interesting finding suggests that consumer decision processes differ in the context of extremely negative and somewhat negative choice sets.

***Study 6: Mediation via the Difference in Perceived Diagnosticity of Average Product Ratings and Review Volumes***

Having established a robust influence of review volume levels on consumers' decisions, we now shift our focus to demonstrating the underlying process by which the effect occurs. Our theorizing suggests that under low review volume levels, consumers acknowledge that their previous evaluations of the average product ratings may not be accurate, and thus, give additional weight to review volumes in their decisions. Conversely, when review volume levels are high, the review volumes merely reinforce the opinion formed of the products, resulting in less weight placed on the review volumes. Thus, in this study we ask participants how important each attribute was to their decision and compute the difference between the importance of average product ratings

and review volumes to demonstrate the changing diagnosticity of these attributes between conditions.

*Participants and design.* One-hundred and eighty-three participants (Amazon mTurk sample; \$0.50 payment) were randomly assigned to one of three review volume levels conditions (low, high, control) in a between-subjects design. The sample size was determined based on balancing a 50-subject per cell rule-of-thumb for online samples at the time the study was conducted with cost considerations. A priori, the goal per cell was 60 subjects.

*Choice set.* Participants saw a choice set of two blenders. Choice options were nearly identical with the exception of their average product ratings and review volumes as described earlier (see Table 1). Similar to Study 2, we used discrete choice as our dependent measure. However, this time we integrated the choice and deferral measures into one, so participants were told that they could choose “Option A”, “Option B”, or “Defer purchase and look elsewhere”. Different from other studies, to test H2, we then asked participants to “indicate the importance of each attribute in making your decision” for the five attributes (image, brand, price, average product rating, review volume) on 7-point scales (1 = not at all important, 7 = extremely important). As expected, there were no significant differences across conditions of the perceived diagnosticity of product image, brand, or price ( $p > .10$ ), because they were relatively comparable across products. Next, we computed a difference score of the review attribute diagnosticities (perceived diagnosticity of average product ratings minus the perceived diagnosticity of review volumes) to demonstrate the changing diagnosticities of the review attributes as a function of the review volume levels. Thus, a positive score indicates that average

product ratings are more diagnostic than review volumes, and vice versa. As the difference score approaches zero, this indicates that consumers would equally weight average product ratings and review volumes in their decisions.

### *Results*

*Choice of the higher-rated, fewer reviews choice option.* Examining participants who chose one of the two product options ( $N = 153$ ), a binary logistic regression, in which we dummy coded our review volume levels, yielded a significant omnibus effect of review volume levels ( $\chi^2(2) = 15.07; p = .001$ ). Consistent with our prior studies, when review volumes were high ( $P_{\text{high}} = 71\%; \chi^2(1) = 11.40; p = .001$ ) or absent ( $P_{\text{control}} = 71\%; \chi^2(1) = 11.83; p = .001$ ) participants were significantly more likely to choose the higher-rated, fewer reviews choice option relative to when review volumes were low ( $P_{\text{low}} = 36\%$ ). There was no significant difference in the high and control conditions ( $p > .95$ ).

*Rate of choice deferral.* A binary logistic regression, in which we dummy coded our review volume levels, yielded a significant omnibus effect of review volume levels ( $\chi^2(2) = 9.42; p = .009$ ). When review volumes were high ( $P_{\text{high}} = 13\%; \chi^2(1) = 4.64; p = .031$ ) or absent ( $P_{\text{control}} = 8\%; \chi^2(1) = 7.60; p = .006$ ), participants were significantly less likely to defer choice relative to when review volumes were low ( $P_{\text{low}} = 29\%$ ), replicating results of Study 3. There was no significant difference in choice deferral between the high and control conditions, consistent with earlier studies ( $p > .40$ ).

*Difference in diagnosticity of review attributes.* A one-way (review volume levels: low, high, control) ANOVA on the difference in diagnosticity of average product ratings and review volumes yielded a marginal omnibus effect ( $F(2, 150) = 4.59; p = .061$ ). Consistent with H2, the difference in perceived diagnosticity between the review

attributes was lower when review volumes were low ( $M_{\text{low}} = .26$ ), relative to high ( $M_{\text{high}} = .85$ ;  $t(150) = 2.28$ ;  $p = .024$ ) or absent ( $M_{\text{absent}} = .75$ ;  $t(150) = 1.89$ ;  $p = .061$ ). There was no significant difference between the absent and high review levels ( $p > .65$ ). In other words, average product ratings were considered significantly more diagnostic than review volumes in the absent and high review volumes conditions, relative to when review volumes were low.

*Preference mediation via the difference in diagnosticity of review attributes.* A mediation analysis (Model 4; Preacher, Rucker, and Hayes 2007) was used to demonstrate that the effect of review volumes (low versus high) on consumer preference was driven by the difference in the diagnosticity of average product ratings and review volumes. As expected, the model demonstrated that the effect of review volumes on consumer preference was mediated via the difference in perceived diagnosticity of average product ratings and review volumes ( $B = -.49$ ;  $CI_{95\%} = [-1.22, -.10]$ ).

### *Discussion*

This study provided support for H2 by demonstrating that the effect of review volumes on choice option preference was driven by the difference in diagnosticity of average product ratings and review volumes: as the perceived diagnosticity of review volumes increases (i.e., when review volumes are low) the preference shifts away from fewer reviews, higher-rated options towards the more reviews, lower-rated option. It further showed, consistent with our propositions, that average product ratings are considered more diagnostic than review volumes, but this difference in diagnosticity is attenuated when review volumes are low. Furthermore, it demonstrated that as the perceived diagnosticity of average product ratings and review volumes become closer,

choice deferral increases, consistent with the findings of Study 2 and prior work demonstrating the link between tradeoff difficulty and choice deferral (Tversky and Shafir 1992; Dhar and Nowlis 1999; Etkin and Ghosh 2017).

### ***Study 7: Mediation using Eye-tracking***

The objective of this study was to demonstrate the robustness of our proposed mediation process by using attentional, rather than self-reported, measures of attribute diagnosticity. We have argued that consumers infer different diagnostic values of average product ratings and review volumes as a function of the level of review volumes. Specifically, in choice sets with neutral and low average product ratings, when consumers see that review volumes are low, it signals to them that average product ratings may not be as diagnostic of product quality as when review volumes are high or when no review volume information is displayed.

In terms of consumers' attention when examining review attributes, we would expect that consumers would be more likely to return to re-examine average product ratings, after viewing low review volumes. This happens, we argue, because consumers need to re-evaluate the average product ratings in light of an additional diagnostic attribute – review volumes. To test this argument, we use eye-tracking measurements to determine not only the gaze times (i.e., time spent looking at) for each attribute, but also the sequence of fixations (i.e., order looked at) for all attributes, to determine whether consumers are more likely to return to re-examine average product ratings after viewing low versus high review volumes.

*Participants and design.* Ninety-two participants (undergraduate sample; course credit) were randomly assigned to one of two review volume levels conditions (low,



high) between-subjects. The sample size was a convenience sample based on the undergraduate participants for a one-week time period. Participants were randomly selected two at a time from a larger sample of research participants to participate in the eye-tracking study. After engaging in a short eye-tracking calibration task, participants followed a similar paradigm to prior studies.

*Choice set.* Participants saw a choice set of two microwaves. Choice options were nearly identical with the exception of their average product ratings and review volumes as described earlier (see Table 1). Relative preference between choice options was measured on the same 7-point scale as in earlier studies.

*Additional measures.* We defined areas of interest (AOIs) as parts of the screen where corresponding product attributes were displayed and measured the number of eye fixations and gaze times for each attribute. Fixations refer to the frequency participants would look at a given attribute, while gaze times refers to the amount of time spent looking at the specific attributes. As expected, there were no significant differences across conditions for fixations or gaze times of product images, brand names, prices, or highlighted information ( $p > .10$ ), so we do not discuss these further.

## *Results*

*Relative preference.* A one-way (review volume levels: low, high) ANOVA on preference yielded a significant effect ( $F(1, 90) = 10.32; p = .002$ ). Consistent with prior studies, preference for the higher-rated, fewer reviews choice option was weaker when review volumes were low ( $M_{\text{low}} = 4.89$ ) relative to high ( $M_{\text{high}} = 3.68$ ).

*Transition matrices.* To provide further support for underlying process, we also derive transition matrices from the eye-tracking data. Doing so allows us to demonstrate

the probabilities of participants transitioning their attention from one attribute to the next. As we discussed earlier, we argue that low review volumes are perceived to be more diagnostic, relative to high review volumes, and this causes consumers to re-evaluate average product ratings. To demonstrate this, we assessed the differential probabilities of participants shifting their attention from review volumes to average product ratings as a function of the review volume levels. Consistent with our theory, participants were significantly more likely to return their attention to the average product ratings after viewing review volumes when the review volumes were low relative to high ( $P_{\text{low}} = .24$ ;  $P_{\text{high}} = .13$ ;  $z = 3.25$ ;  $p < .01$ ). This suggests that participants were nearly twice as likely to return their attention to average product ratings when review volumes were low versus high. Importantly, the transition proportions from review volumes to all other attributes were not significantly different across conditions ( $p > .10$ ).

-----

Insert table 3 about here

-----

*Difference in fixation counts for the review attributes.* Because our variable of interest is the difference in attention paid to average product ratings and review volumes, we calculated the difference in fixations between average product ratings and review volumes. A one-way (review volume levels: low, high) ANOVA on the difference in fixation counts yielded a significant effect ( $F(1, 90) = 7.05$ ;  $p = .009$ ). As expected, the difference in fixations between average product ratings and review volumes was greater

when review volumes were low ( $M_{\text{low}} = 4.29$  fixations) relative to high ( $M_{\text{high}} = 2.11$  fixations). This is consistent with our view that low review volumes cause consumers to re-evaluate the average product ratings attribute, thus increasing overall attention paid to that attribute.

*Difference in gaze times for the review attributes.* A one-way (review volume levels: low, high) ANOVA on the difference in gaze times yielded a significant effect ( $F(1, 90) = 10.59; p = .002$ ). As expected, the difference in gaze times between average product ratings and review volumes was greater when review volumes were low ( $M_{\text{low}} = 12.40$  seconds) relative to high ( $M_{\text{high}} = 5.78$  seconds). Consistent with our prior finding, consumers seem to pay more attention to average product ratings when review volumes are low versus high, and we argue that this occurs because the low review volumes cause consumers to re-evaluate their inferences from the average product ratings.

*Mediation via the difference in gaze times.* We argue that gaze times are a more precise measure of attention relative to fixations because they quantify the time spent on an attribute. As such, we demonstrated that consumers are likely to pay more attention to average product ratings when review volumes are low, and thus, this difference in gaze times would mediate the influence of review volumes on consumer preference. Using the mediation analysis (model 4; Preacher and Hayes 2007), we demonstrated that the difference in gaze times between average product ratings and review volumes mediated the effect of review volumes on consumer preference between choice options ( $B = .25$ ;  $CI_{95\%} = [.04, .62]$ ).

*Discussion*

This study provided further evidence for H2, by demonstrating that the difference in gaze times for the average product ratings and review volumes mediated consumer preference between options. When review volumes are low, it signals to consumers that the average product ratings may be relatively closer in diagnosticity to review volumes, leading them to re-evaluate this attribute before reaching a decision. Yet, when review volumes are high, there is a hierarchy in diagnosticity between two attributes, and consumers can reach a decision faster without re-evaluating average product ratings.

## **GENERAL DISCUSSION**

Across seven studies we find consistent support for our propositions that average review ratings are more diagnostic cue of product quality than review volumes and that review volumes can become more influential in consumer decisions when: a) average product ratings are low or neutral and b) review volume of the choice set is low. This change in attribute diagnosticity leads to a systematic shift in preference away from the higher-rated, fewer reviews option towards the lower-rated, more reviews option.

Furthermore, when the diagnosticity of these two review attributes are closest to each other, consumers experience tradeoff difficulty as evidenced by increased choice deferral.

### *Robustness Checks and Potential Moderators*

Appendix B, contains eight additional experiments that rule out alternative explanations and test additional potential moderators of the review volume effects demonstrated in this paper. The studies are reported in full in Appendix B and outlined briefly next.

In Appendix B2, we examine the role of relative versus absolute differences in review volumes. The literature on numerosity suggests that viewing attributes on expanded versus contracted scales (e.g., months versus years) changes consumers' interpretations of the attribute, as the absolute difference in values may change while keeping the relative difference constant (Burson, Larrick, Lynch 2009; Pandelaere, Briers, and Lembregts 2011; Lembregts and Pandelaere 2013). Our work differs from this for two reasons. First, our focus is on the effect of an attribute when the absolute difference is held constant rather than the relative difference. This is important as it tests a best-case scenario for the underdog in which there is no "first mover's advantage" in the acquisition of future reviews. Second, low versus high review volumes are not direct comparisons to contracted versus expanded scales as there is diagnostic value in the actual values provided. However, it is possible to draw a hypothesis from the findings of this literature in suggesting that absolute difference in attributes is more influential than the relative difference. In addition to the low and high review volume levels used as in prior studies (e.g., 12 versus 45 and 212 versus 245), we also included a high review volume level condition where the relative difference was held constant with the low condition (e.g., 212 versus 795). We replicate the findings of our previous studies in the low and high conditions, while also demonstrating an effect consistent with the numerosity literature for the high-relative condition. Yet, in comparing the effect sizes, we find that our effect is stronger for multiple product categories.

In Appendix B3, we compare the diagnostic value of the aggregate reviews ratings with more disaggregated information, ratings distributions the ratings of individual reviews. Some literature investigating the role of a ratings distribution has

demonstrated a significant effect of skew on product evaluations (Fisher, Newman and Dhar 2018; Khare et al. 2011). This work largely differs from ours in that it explores only single-option choice sets. In the context of multi-option choice sets, we argue that it is more cognitively-demanding to compare and interpret different ratings distributions relative to the other review attributes (i.e., average product ratings and review volumes). This is because the ratings distribution is a more complex attribute with several values, one for each rating. Thus, we argue, and demonstrate that the effect of a ratings distribution is diminished in multi-option choice sets. We tested this effect using several different distributions (e.g., positively-, negatively-, and even-skewed).

In Appendix B4, we further examine whether aggregate reviews information is diagnostic in the presence of other diagnostic information. Specifically, we look at the context of products with primarily aesthetic value. Prior literature has suggested that consumer responses to reviews differs as a function of the self-expression goals of the consumer and the product being evaluated (He & Bond 2015; Rozenkrants, Wheeler, and Shiv 2017). Consistent with this literature, in this study, we demonstrate that aggregate review attributes are less diagnostic when consumers can easily infer preference based on the aesthetics of the good, rather than concerns of functional attributes. Thus, we expect and demonstrate that the effects of review volumes are attenuated when consumers face choice sets with high aesthetic value and low functional value (e.g., artwork).

In Appendix B5, we examine the role of single versus joint evaluations. Prior literature investigating this aspect of choice has demonstrated that joint evaluation increases the diagnosticity of difficult-to-evaluate attributes (Hsee 1996) and attenuates the effects of numerosity (Schley, Lembregts, and Peters 2017). In our research, we argue

that average product ratings and review volumes are both common attributes to consumers which allows them to easily hold values in memory across choices. As such, single versus joint evaluations would play a minimal role in choice sets where average product ratings and review volumes are the differentiating attributes. Thus, we demonstrate a null effect of evaluation mode in Appendix B5.

In Appendix B6, we examine the role of the presence versus absence of a popularity cue. Some people might theorize that review volumes act solely as a signal of product popularity. If this were the case, the effect of review volumes should be attenuated in the presence of another popularity cue. Thus, in this study we test the effect of labeling options as a “Best Seller” and demonstrate that this does not attenuate the effect of review volumes. We argue that this occurs because it is more challenging to quantify the “Best Seller” label relative to review volumes, thus it is difficult to determine the diagnostic value of the label.

In Appendix B7, we examine the role of production years of the goods. One reason that some products have more reviews than others is simply because they have been on the market longer. While this may be a positive signal for some products, for tech products, this should signal outdated technology. We argue that the effect of review volumes is so strong, it leads consumers to discount the inferior technology if it has more reviews. We demonstrate that consumers are more apt to choose older products with more reviews (e.g., a 2013 DVD player or Galaxy S6) relative to newer products with fewer reviews (e.g., a 2015 DVD player or Galaxy S7) when the level of review volumes is low versus high. As such, it is quite likely that review volumes may bias consumers’ decisions, leading to inferior, or suboptimal, outcomes.

In Appendix B8, we examine the role of a “new arrival” label on the effect of review volumes. Because new products have been on the market for a shorter period than other products, a lower review volume should be justified. In theory, consumers should be more accepting of “new arrivals” with low review volumes relative to options that do not feature this label as their low volume may not be justified. However, it is difficult to quantify the diagnostic value of a “new arrival” label and arrive at a proper discount rate for the review volumes relative to simply comparing the review volumes of competing products. As such, we theorize, and demonstrate that the effect of review volumes persists in the presence of a “new arrival” label.

In Appendix B9, we examine the role of credible reviews. Some may argue that low review volumes contain more risk relative to high review volumes due to a higher likelihood of fraud (Luca and Zervas 2016; Mayzlin, Dover, and Chevalier 2014). Thus one might conclude that if consumers could be certain of the veracity of low review volumes, it may attenuate the effect of review volumes. Yet, once again, we argue that it is cognitively-demanding to interpret and quantify the value of “credible” reviews relative general reviews, resulting in a null effect. We test this proposition by labeling the higher-rated, fewer reviews option as “Consumer Reports Verified” and demonstrate once again, that this label does not attenuate the effect of review volumes.

Across the eight experiments featured in Appendix B, we consistently demonstrate the persistent effect of review volumes. While it is quite plausible that additional moderators of review volume effects, besides those exhibited in this paper, exist, it seems that many of the additional attributes websites use to help their consumers do not actually attenuate the weight consumers place on review volumes. As such,



retailers should be cautious in how they present review volumes to consumers as it is a very powerful attribute in shaping decisions.

### *Managerial Implications*

Next, we provide an illustration of how the results of our study can be incorporated into the business practices. We interviewed two managers involved with product review acquisition strategies for their respective brands. One, the CEO of a nutritional supplement company, instructs his team to employ a proactive strategy where they aggressively pursue reviews from customers via email nudges after purchase, as well as steep product discounts. The latter strategy is used to increase the absolute number of reviews as a function of more sales, whereas the former strategy is used to increase the sales-to-review conversion ratio. The other, an associate brand manager for a leading home electrics company, relies on a reactive strategy in which they offer free product in exchange for honest reviews once they perceive their sales to have stagnated. This strategy drives review volumes but not sales. Both managers mentioned that product ratings and review volumes are very important for their sales, and they would like both attributes to excel. When pressed on the tradeoff of the two attributes, both managers said that they would prefer a 4.0 rated product with 100 reviews over a 4.3 rated product with 15 reviews. Furthermore, when we discussed that the promotions may bring about heterogeneity in customers (i.e., customers who normally would not use the product, and therefore, be more likely to have less of a fit with the product), they both agreed that they would welcome customers who did not love their product in exchange for a higher review volume, assuming that it did not bring a wave of completely negative reviews. Next, we will discuss how managers could benefit from actively managing their review volumes to

increase review volumes in early- and mid-stage product life cycles when they have competitors with more reviews.

We conducted a simulation where the higher-rated good with fewer reviews accrued reviews at the same rate as the lower-rated, more reviews good or at twice the rate (e.g., when a review acquisition strategy was employed). We then crossed this with whether the firm employed a proactive review acquisition strategy early on in the product life cycle (e.g., at 5 reviews) or a reactive strategy later on in the life cycle (e.g., at 15 reviews). The simulation was conducted in MATLAB using choice shares computed from Study 1. We took the relative preference measure used in that study and pooled the numbers that indicated preference for the higher-rated, fewer reviews option (i.e., “1 – 3”) and the numbers that indicated preference for the lower-rated, more reviews option (i.e., “5 – 7”). Participants who indicated no preference (i.e., “4”) were excluded from the choice share computation. This resulted in several data points ranging from low to mid to high review volumes. Then we used the MATLAB surface fitting function (“fit”) to interpolate choice shares for various combinations of review volume pairs.

Our simulations were agnostic as to what type of review acquisition strategy was employed (e.g., email nudges, price promotions, free product, etc.), but we modeled the effect of employing an acquisition strategy by doubling the likelihood of a review being written. We seeded the baseline likelihood of receiving a review after a purchase at 50%, and the likelihood of receiving a review when a review acquisition strategy is employed at 100%. The likelihood of leaving a review has been estimated as low as .001% depending on the product category. Thus, the results we discuss here are extremely

conservative estimates of the impact review volumes might have on choice shares over time.

In the first two simulations, we investigated a proactive strategy where the higher-rated option had five reviews, whereas the lower-rated option had 10 reviews. In Scenario A, we held the likelihood of receiving a review constant across both products at 50%. Thus, the first scenario investigated how long it would take for a product with a higher rating, but five fewer reviews, to accumulate more sales than the other products. As seen in figure 8A, we see that it would take approximately 240 consumers purchasing options in this category. In Scenario B, we assume that the manager of the higher-rated product employs an active review management strategy and doubles the likelihood of receiving product reviews. In this scenario, we see that the higher-rated product reaches the dominant sales position roughly 33% more quickly, or around 160 consumer category purchases (see figure 8B).

-----

Insert figures 8A-D about here

-----

In Scenarios C and D, we investigate a reactive strategy where the market has matured (i.e., 15 and 45 reviews, respectively). Once again, in Scenario C we assume that each product has a 50% likelihood of receiving a review after it is purchased. Here, it takes about 75 additional customer purchases in the product category before the higher-rated good surpasses the lower-rated good in sales (see figure 8C). In Scenario #4, we

once again assume that the higher-rated good actively manages their reviews, and can double the likelihood of receiving a review after purchase. This results in the higher-rated good surpassing the lower-rated good in sales roughly 20% earlier, or after 60 consumer category purchases (see figure 8D).

While these scenarios are just a few of the many possible that exist in marketplace, they provide additional support for our claim that review volumes are highly influential in consumer decisions, and managers would be wise to oversee their growth. As new products with fewer reviews enter against incumbents with more reviews, employing a proactive strategy to spur review volume growth can quickly decrease the disadvantage that a product manager faces when competing against established products. While there may be costs associated with discounted products or additional review nudging, early on in the product life cycle, it would appear to be quite beneficial to incur these costs.

### *Conclusions and Future Research*

This research outlines conditions where the diagnosticity of review volumes as cues of product quality increases relative to diagnosticity of average product ratings, potentially leading to suboptimal decisions for consumers, and an increase in choice deferral for brands and retailers. Theoretically, we argue that an inherent difference in the types of scale in which these attributes are presented (bound and unbound) leads to the observed difference in their diagnosticity, and by demonstrating how consumers integrate attributes on both scales into a single judgment, we contribute to the literature in numerical cognition. Furthermore, we provide some clarity to the debate on the relative influence of average product ratings and review volumes (see: Floyd et al. 2014; You,

Vadakkepatt, and Joshi 2015) by demonstrating the conditional influences of volume under various valence conditions.

Future research should explore how review volumes affect interpretation of average product ratings presented on expanded (e.g., 0 – 100%) or more contracted bound scales (e.g., “Thumbs Up/Down” votes). Furthermore, research on strategies to bind review volumes to a scale (e.g., classifying review volumes based on “ideal” or “sufficient” values but not quantifying the actual number) can attenuate the bias created by review volumes. Future research could also explore how the relative versus absolute difference in values across the attributes would affect the difficulty of tradeoffs between the review attributes.

From a managerial perspective, future research demonstrating how consumers interpret the presence versus absence of different product attributes across multiple retailers would be an interesting avenue, though this is ultimately an empirical question to answer. Furthermore, in applying these findings in a field setting, individual retailers could determine the optimal strategy for their website. For example, given the demographics and preferences of a specific retailer’s clientele, would their customer base prefer one strategy over another? Lastly, while consumers read very few reviews before making decisions (BrightLocal 2017), exploring the tradeoffs made between review content and aggregate review valence and volume is an important future direction. While the aggregate information is more representative of the products, individual reviews may heighten saliency of specific product details, leading to shift in weights of the various attributes in their decisions. For example, a review highlighting a bad service experience may outweigh the aggregate information indicating that a restaurant is rather popular with

high quality food. Thus, understanding when consumers want to read reviews in addition to summary information, and the relationship between these two sources of information is also an important future direction.

CHAPTER III: ESSAY II

Red Flag! The Consequences of Alerting Consumers  
to Fake Reviews

Jared Watson, Dr. Amna Kirmani

*“TripAdvisor has reasonable cause to believe that individuals or entities associated with or having an interest in this property may have interfered with traveler reviews.” –*

*TripAdvisor, 2018*

Defined as those written with the intent to mislead or deceive, fake reviews appear to be a major concern online. Recent research suggests that upwards of 15% of online reviews may be fake (Luca and Zervas 2016). And this seems to be consistent with industry perceptions, as Yelp has disclosed that they remove roughly 20% of their reviews for lacking credibility and helpfulness. Amazon has even taken legal action in an attempt to curb the problem, recently suing over 1,100 individuals they believe to have created fake reviews (Tuttle 2015). Yet, the lawsuits may not have made much impact as a recent Forbes article concluded that “Amazon’s fake review problem is now worse than ever” (Woolacott 2017). Indeed, an entire industry has arisen around the creation of fake reviews, as “online reputation management” firms will post positive reviews for your business, and negative reviews for your competitors, for a nominal fee (Segal 2011). This would suggest that the fake review problem may be more pervasive than academics have previously considered.

In an effort to increase transparency, some websites (e.g., Yelp and TripAdvisor) inform consumers when they have caught a brand featured on their website involved in the creation of fake reviews, but it is unclear how consumers respond to this message. We term this message a fake review alert. For example, once Yelp identifies that a brand has attempted review fraud (e.g., purchasing fake reviews, incentivizing positive reviews, etc.), that brand’s sub-page within Yelp will feature an alert for the next 90 days (Yelp 2013; Curtis 2014). These alerts, like the one above, explain the brand’s infraction, but



they are only temporary. Do consumers penalize the brand by lowering ratings? Does their reading behavior for reviews change? We address these questions in this paper.

Prior research has examined the characteristics of fake reviews, as well as the brands which are likely to solicit them. Computer scientists have developed complex algorithms which determine the veracity of reviews based on certain syntactical cues (e.g., excessive capitalization or repetitious punctuation) and contextual cues (e.g., IP addresses, review frequency, etc.) (Feng, Banerjee, and Choi 2012; Akoglu, Chandy, and Faloutsos 2013; Ott et al. 2013; Mukherjee et al. 2013). In marketing, a few papers have investigated the brand characteristics of firms that solicit fake reviews and the propensity for a brand to solicit fake reviews. Mayzlin, Dover, and Chevalier (2014) find that independent brands have a greater incentive to solicit fake reviews relative to chain brands, and competitors of independent brands are likely to leave fake negative reviews, when customers are not verified. Similarly, Luca and Zervas (2016) find that a restaurant is more likely to engage in review fraud when it has a weak reputation (e.g., few reviews or recent negative reviews), and that chain restaurants are less likely to engage in review fraud relative to independent restaurants, presumably because the brand name is better known.

Rather than identifying characteristics of fake reviews or the brands which solicit them, our paper investigates the effects of fake review alerts on how consumers process reviews and, ultimately, affect their brand ratings. In general, consumers read reviews in order to make an informed decision (Gilbert 1991). But in the presence of a fake review alert, consumers become suspicious of the information, leading to concerns of information accuracy. We propose that the effects of a fake review alert depend on

whether the alert is for positive or negative reviews. A fake positive (negative) review alert discloses that reviews which have attempted to increase (decrease) brand evaluations have been identified. In conjunction with accuracy concerns, a fake review alert activates persuasion knowledge, leading consumers to be suspicious about the brand's other reviews. As a result, they are more likely to read negative reviews in order to avoid making a poor decision.

In terms of brand ratings, we propose that a fake review alert will lead consumers to deviate from the average rating in an attempt to correct for false information. When persuasion knowledge is activated, consumers adjust their initial impressions to correct for the perceived bias inherent in the persuasion attempt (Kirmani and Campbell 2004). This correction means that consumers are likely to increase the brand rating in the presence of a fake positive review alert and decrease the rating in the presence of a fake negative review alert. At the same time, because a fake review indicate that someone is trying to mislead them, consumers may use their brand ratings to seek justice against the source of the fake review.

We use multiple methodologies to investigate the effect of a fake review alert on consumer review search and brand ratings. In Study 1, we employ web-scraping to collect data on brands that have received fake positive review alerts, and analyze their brand ratings before, during, and after the alert is active. We demonstrate that brand ratings decrease during the alert but are restored after the alert is removed. In Study 2, a longitudinal laboratory experiment replicates these findings and shows a different pattern for fake negative review alerts; brand ratings increase during the alert, and this positive increase persists after the alert is removed. Study 3 then demonstrates the process by

which perceptions of average rating accuracy and the desire for justice mediate the effect of a fake review alert on brand ratings. Finally, Study 4 demonstrates how consumers' overcorrection in the presence of a fake review alert can be attenuated.

In the following sections, we develop our conceptual framework regarding consumers' responses to fake review alerts and propose several hypotheses. We then test these hypotheses employing multiple methodologies and conclude with a discussion of the implications and directions for future research.

## **CONCEPTUAL BACKGROUND AND HYPOTHESES**

Reviews affect consumers' product evaluations (Khare, Labrecque, and Asare 2011; Moe and Trusov 2011; Schlosser 2011; Chen and Kirmani 2015) and purchase incidence (Chevalier and Mayzlin 2006; Clemons, Gao, and Hitt 2006; Dellarocas, Zhang, and Awad 2007; Duan, Gu, and Whinston 2008; Chintagunta, Gopinath, and Venkataraman 2010; Zhu and Zhang 2010; Sun 2012; Ho-Dac, Carson, and Moore 2013; Floyd et al. 2014; You, Vadakkepatt, and Joshi 2015). Almost 85% of consumers now trust online reviews as much as a personal recommendation (BrightLocal 2017), because reviews can reduce uncertainty and transaction costs (Brown and Reingen 1987; Murray 1991; Banerjee 1992; Udo and Marquis 2002). In general, online reviews are perceived as a source of diagnostic information rather than a source of persuasion. Thus, in the absence of a fake review alert, consumers are likely to form opinions consistent with the information they encounter. This means that prior to experiencing the brand, consumers' ratings are likely to be consistent with the brands' average rating.

We first discuss how consumers might process reviews in the presence of a fake review alert as well as the effect on brand ratings. We then discuss the mediation by the desire for justice and accuracy perceptions. Lastly, we discuss how the effect of a fake review alert can be attenuated by reinforcing the aggregate brand information after consumers encounter an alert.

### **Review Processing in the Presence of a Fake Review Alert**

A fake review alert informs consumers that someone has attempted to present them with fraudulent information. Once fake reviews are identified, websites such as Yelp remove the fake reviews, and post an alert. While this suggests that the remaining reviews are trustworthy, the alert creates suspicion regarding the veracity of the available information (Kirmani and Zhu 2007). We argue that the fake review alert leads to the activation of persuasion knowledge (Friestad and Wright 1994), as consumers draw inferences about the source and intention of the fake reviews. Although fake review alerts inform consumers that misleading information has been identified and removed, it increases the saliency of fake reviews, leading to an attempt to correct bias.

#### **Persuasion Knowledge and Correction**

Persuasion knowledge is activated when consumers are suspicious of a marketing agent's motives, statements, or actions, which results in less favorable brand and agent attitudes (Campbell 1995; Campbell and Kirmani 2000; Jain and Posavac 2004; Kirmani and Zhu 2007). For example, when consumers are suspicious of an advertising tactic, they may decrease their purchase intentions for the focal product (Campbell 1995). Sometimes, however, accessing persuasion knowledge can lead to increased brand and agent attitudes (Isaac and Grayson 2017). For example, when a marketer employs a

credible tactic (e.g., reinforcing the availability of a price-matching opportunity), it can actually increase perceptions of the marketer and purchase intentions. In our research, we argue that consumers perceive fake reviews as manipulative, leading to negative effects.

When persuasion knowledge is activated, consumers correct their initial impressions to take into account the information they perceive to be biased (Campbell and Kirmani 2000). This means that consumers will determine the valence and size of the bias, and adjust their brand perceptions. However, the flexible correction model (Petty and Wegener 1993; Wegener and Petty 1995b; Wegener and Petty 1997) suggests that people can easily overcorrect due to inaccurate perceptions of the bias. In the context of a fake review alert, consumers are unlikely to account for the trustworthy review volume due to the salience of fake reviews, leading to brand rating corrections even when there are potentially hundreds of accurate reviews.

The type of correction will depend on the valence of the fake review alert. Fake positive reviews are created by companies with the goal of increasing favorable perceptions of their brand in an attempt to attract customers. Companies also incentivize consumers to write these reviews in exchange for discounts. In contrast, fake negative reviews are created with the goal of increasing unfavorable perceptions of the brand. The likely source of these reviews is a competitor of the brand, who may also pay third parties to write the negative reviews (Mayzlin, Dover, and Chevalier 2014). Thus, the source of fake positive reviews is the brand, whereas the source of fake negative reviews is the competition.

Following the work of Wegener and Petty, we argue that consumers use a naïve theory of persuasion in an attempt to correct for bias (Thompson 1981; Baumeister and

Newman 1994). After determining the valence of the bias, consumers will adjust their brand ratings in the opposite direction of the perceived bias without properly accounting for the size of the bias. Because fake review alerts signal that the fake reviews have been identified and removed, only reviews presumed to be authentic by the website remain. The authentic reviews are what factor into the average brand rating and review volume, but consumers may still correct for a perceived bias. Formally:

**H1:** *Relative to when a fake review alert is absent, a fake negative (positive) review alert increases (decreases) brand ratings.*

### **Mediation via Perceived Rating Accuracy and the Desire for Justice**

Because reviews reduce uncertainty in consumers' decisions (Brown and Reingen 1987; Udo and Marquis 2002), consumer's brand ratings are dependent on how accurate they perceive the reviews to be. Thus, the expectant brand rating is a function of the perceived average rating accuracy. And relative to when an alert is absent, a fake review alert decreases the perception of average rating accuracy, regardless of valence.

Furthermore, the desire for justice may also come into play for positive fake review alerts. Justice is a multi-faceted construct relating to the outcomes, interactional behaviors, and procedures that arise in a situation (Tax, Brown, and Chandrashekar 1998). Literature exploring the desire for revenge has associated it with intense cognitive (McCullough, Fincham, and Tsang 2003) and emotional (Bonifield and Cole 2007) investment. If a brand mistreats a consumer, that consumer may seek justice in the form of revenge by spreading negative word of mouth about their experience (Ward and Ostrom 2006) or avoiding the brand (Bechwati and Morrin 2003; Bonifield and Cole 2007; Ward and Ostrom 2006; Grégoire, Tripp, and Legoux 2009). While a desire for

revenge may persist as long as consumers hold a grudge (McCullough, Fincham, and Tsang 2003), it is likely to fade over time (Bies, Tripp, and Kramer 1997) due to the psychological cost associated with maintaining the desire (Bechwati and Morrin 2003). Consumers can further satisfy their desire for revenge by punishing the brand at-fault (Grégoire, Tripp and Legoux 2009).

However, consumers will only punish the brand when they can attribute the fake reviews to the brand itself. Since the positive fake review alert states that the brand is the source of the fake reviews, consumers are likely to blame the brand for trying to mislead them. Thus, the desire for justice will be stronger in the presence of a fake positive versus negative review alert.

**H2:** *A fake review alert, regardless of valence, decreases the perception of average rating accuracy.*

**H3:** *Relative to when a fake review alert is absent or negative, a fake positive review alert increases the desire for justice.*

Taken together, we argue that brand ratings are mediated by the perception of average rating accuracy in the presence of a fake negative review alert, while being mediated by the perception of average rating accuracy and the desire for justice in the presence of a fake positive review alert. Formally:

**H4a:** *In the presence of a fake negative review alert, the effect of the alert on brand ratings is mediated by the perception of average rating accuracy.*

**H4b:** *In the presence of a fake positive review alert, the effect of the alert on brand ratings is mediated by both the perception of average rating accuracy and the desire for justice simultaneously.*

### Effects of Fake Review Alerts on Reading Behavior

The correction process entails not only a deviation from initial brand impressions but also changes to how consumers acquire information (Wilson and Brekke 1994; Wegener and Petty 1995b; Wegener and Petty 1997). So the fake review alerts have the potential to bias consumers' decisions by not only leading them to correct for a bias that has already been corrected for by the website, but also by leading them to read reviews which they suspect are more credible and avoid reviews which they suspect are less credible.

One benefit of online reviews is that they generally provide consumers access to a variety of opinions (Udo and Marquis 2002), so consumers can find both negative and positive reviews. Because the alert makes consumers suspicious, they may be more likely to scrutinize negative rather than positive reviews. Since the alert makes them question the accuracy of the reviews, they may want to avoid making a poor decision. This suggests that they will pay more attention to negative than positive information. The negativity bias suggests that consumers attend to, and elaborate more on, negative information (Baumeister et al. 2001; Rozin and Royzman 2001). It has been demonstrated in the context of online reviews as well (Schlosser 2005). Specifically, Schlosser found that consumers' reviews were more influenced by others' negative reviews, relative to positive reviews. As such, we would expect that consumers generally read more negative than positive reviews in an effort to learn what brands to avoid.

Formally:

**H5:** *The average valence of reviews read by consumers will decrease in the presence of a fake positive (versus negative) review alert.*



## **Moderating the Effect of Fake Review Alerts on Brand Ratings**

Our hypothesized process is that consumers' brand ratings are a function of their perceived accuracy of the average rating, as well as their desire for justice in the presence of a fake positive review alert. One implication of our theoretical process is that making the average rating salient should attenuate the effect of a fake review alert on a brand's rating. Drawing attention to the review volume should make consumers perceive the reviews as more accurate and lessen the impact of the few fake reviews. Formally:

**H6:** *Presenting a brand's aggregate information after a fake review alert is displayed attenuates the desire for justice, and ultimately, attenuates the effect of the fake review alert on brand ratings.*

In the next section, we test these hypotheses over several studies.

## **OVERVIEW OF STUDIES**

We present four studies to test the hypotheses. Study 1 uses web-scraped data from Yelp.com to provide support for the effect of a fake positive review alert on brand ratings. The next three studies were conducted in the lab. Study 2 tests the effect of fake positive and negative review alerts on brand ratings (H1), the perception of average rating accuracy (H2), the desire for justice (H3), and the average valence of reviews read (H5) over three rounds. Study 3 then tests all of these hypotheses in a single period, while also testing H4 by demonstrating the dual mediating roles of the perceptions of average rating accuracy and the desire for justice. Lastly, Study 4 provides evidence for H6 by

demonstrating the moderating role of an aggregate information disclaimer on the effect of fake review alerts, while also providing additional evidence for H1 – H5.

-----

Insert table 1 about here

-----

### **Study 1: Empirical Investigation of Positive Fake Review Alerts on Yelp**

Study 1 examines the effect of fake positive review alerts on consumers' brand ratings on Yelp.com. In 2012, Yelp began their "Consumer Alert" program in which brands caught attempting to manipulate reviews received an alert displayed on their subpage within Yelp for 90 days. This alert informs consumers who visited the brand's subpage that the brand had been caught attempting to manipulate reviews in an attempt to inflate the rating of their business. After 90 days, the alert is removed. Yelp currently displays alerts for brands which have been caught trying to increase their own rating, but not for those who have received fake negative reviews from competitors.

In addition to the fake review alerts, Yelp uses an algorithm to determine the likely veracity of reviews and automatically filters reviews into "recommended" and "not recommended" lists. Not recommended reviews are separated from the recommended reviews, are not prominently displayed, and do not impact the calculation of the average rating and review volume. Although Yelp's sorting algorithm is proprietary, it is thought to include attributes such as the reviewer's IP address, the reviewer's review volume, and syntactical cues within the review. While reviews can be filtered for reasons other than

being fraudulent (e.g., lacking usefulness), they do serve as a proxy for fake reviews. Thus, we examine the effect of a fake review alert on both the recommended and not recommended reviews

### *Data and Procedure*

To create the dataset, we analyzed news articles and press releases that mention the “Consumer Alert” program, which resulted in a list of 32 brands across 11 product categories (e.g., medical, restaurants, entertainment, etc.) that were known to have received the fake review alert. We then web-scraped all Yelp reviews of these brands. This resulted in an initial dataset of 32 brands and over 9000 reviews. However, to avoid concerns of a brand’s time on the market, we trimmed the data for each brand to only include a 270-day timeframe split into three periods: 90 days *before* the alert was active, 90 days *during* the active alert, and 90 days *after* the alert. This resulted in 1885 reviews for the 32 brands. These reviews were further classified as “recommended” and “not recommended” reviews by Yelp’s algorithm, which served as a proxy for authentic and fake reviews. Though tangential to the purpose of this research, this allowed us to examine the effect of a fake positive review alert on both review types. The time periods and review types were both treated as between-subjects factors, as we assume that most consumers are not leaving multiple reviews for the same brand. The dependent variable of interest was the average rating, so we extracted the rating from each review, and pooled the data across brands. We used the average ratings that occurred in the 90 days before the alert as the baseline, and compared the ratings during and after the alert to determine the effect of a fake positive review alert.

### *Results and Discussion*

H1 posits that relative to when a fake review alert is absent, a fake positive review alert decreases brand ratings. A 2 (review type: not recommended, recommended) x 3 (90-day time period: before alert, during alert, after alert) ANOVA on brand ratings yielded main effects of review type ( $F(1, 1879) = 18.35; p < .001$ ) and time period ( $F(2, 1879) = 94.83; p < .001$ ), qualified by a significant interaction ( $F(2, 1879) = 4.18; p = .015$ ). *Before* the alert, ratings were significantly higher for the not recommended than the recommended reviews ( $M_{\text{not recommended}} = 4.64, M_{\text{recommended}} = 4.01; F(1, 1879) = 59.74; p < .001$ ). This finding is consistent with the notion that not recommended reviews contained more fake positive versus negative reviews. *During* the alert, ratings were marginally higher for the not recommended reviews ( $M_{\text{not recommended}} = 3.26, M_{\text{recommended}} = 2.98; F(1, 1879) = 3.29; p = .07$ ). We define a p-value as marginally significant based on the rule-of-thumb that it falls between .05 and .10. This suggests that an alert attenuates the prevalence of fake positive reviews. After the alert, ratings were not significantly different for review types ( $M_{\text{not recommended}} = 4.19, M_{\text{recommended}} = 4.04; F(1, 1879) = .76; p > .35$ ), which suggests that fake review alerts may have a long-lasting impact in reducing the volume of fake reviews. The main effects demonstrated that the ratings of not recommended reviews were higher than those which were recommended ( $M_{\text{not recommended}} = 4.35, M_{\text{recommended}} = 3.81$ ), while the ratings before an alert ( $M_{\text{before}} = 4.33$ ) were greater than those after an alert ( $M_{\text{after}} = 4.12; t(1882) = -2.63; p = .009$ ), which were greater than those during an alert ( $M_{\text{during}} = 3.13; t(1882) = -8.33; p < .001$ ).

These findings provide initial evidence for H1, as the presence of a fake positive review alert was shown to decrease brand ratings. Although unlikely, it is possible that the changes in brand ratings corresponded to changes in actual quality, or unique to the

consumers who left reviews in each period. Thus, in the following studies, we use controlled lab studies to avoid these concerns. While outside the scope of this research, this study also provided evidence for the long-term effects of a fake review alert. It appears that the effect of the fake positive review alert on brand ratings only occur so long as the alert active, but may curb fake reviews in the long-term, as witnessed by the lower rating of the “not recommended” reviews. While this research focuses on the effects of alerts on consumers, rather than fake review providers, it does provide some interesting insights into firm behavior as well.

-----

Insert figure 9 about here

-----

## **Study 2: The Effect of Fake Negative Review Alerts**

Study 2 provides tests of H1 – H3 (the effect of fake review alerts on brand ratings, perceptions of average rating accuracy, and the desire for justice), and also H5 (the effect of fake review alerts on reading behavior). We employed a longitudinal experiment that took place over three rounds of data collection. The manipulation occurred in the second round when participants saw either a fake negative or positive review alert. In the first and third time periods, fake review alerts were not present. Thus, we employed a 2 (fake review alert: negative, positive) x 3 (time period: before alert, during alert, after alert) mixed design, where the alert was between-subjects and the time period was within-subject. Therefore, it is important to note that this study examines the

effect of a fake review alert on the same reviewer's brand ratings over time, whereas Study 1 assumed that the reviewers were unique in each time period.

### *Design and Procedure*

Participants were undergraduates at a large American university who participated in three rounds of studies in exchange for course credit. Round 1 included 388 participants. Round 2 included 375 participants. Round 3 included 348 participants. Ultimately, 278 participants (48% female;  $M_{\text{age}} = 20.38$ ) completed all three rounds, representing an attrition rate of 24.93%. In Round 1, all participants completed the fake review alert absent condition. In Round 2, participants were randomly assigned to either the fake positive or negative review alert condition. In Round 3, all participants completed a fake review alert absent condition. Thus, the between-subjects manipulation only occurred in Round 2. The sample size was a convenience sample based on the undergraduate participants for a two-week time period in each round. In Round 1, participants read a scenario which detailed an upcoming trip to San Diego, CA, which was far away from their university. They were told that they had not yet booked a hotel, and thus, decided to view some hotel reviews online. In both of the subsequent rounds, they were told that they still had not booked the hotel and decided to read some additional reviews. Using a template similar to Yelp, at the top of the page participants viewed summary information of the hotel detailing the location (San Diego, CA), price range (\$\$), average rating (3.2 out of 5 stars), number of reviews (264), and a few photos of the hotel room (see Appendix C). They also saw 11 snippets of reviews (ten labeled “recommended” and one labeled “not recommended”). The “not recommended” review was included to determine if participants were interested in reading information that was

likely fake. The review set contained two reviews of each valence (1-star, 2-star, etc.) plus the “not recommended” review. Furthermore, the not recommended review had a disclosure informing participants that this review lacked credibility and did not impact the average rating or number of reviews. After viewing the review snippets, participants had the opportunity to read the full-length review for any, and as many, of the 11 reviews as they wanted, one at a time.

After reading the reviews, participants were asked a series of questions. The *brand rating* was assessed by asking participants “What star rating would you assign this brand? That is, what rating do you think reflects the true quality of West End Hotel?” (1 – 5 stars, continuous scale). *Perception of average rating accuracy* was assessed with the question “Given the available information, how accurate is the 3.2 out of 5.0 average star rating from other consumers for West End Hotel?” (1 = not at all – 7 = extremely). The *desire for justice* was computed by calculating a difference score of two measures: “To what degree would you say that your rating was made to punish the brand” and “To what degree would you say that your rating was made to reward the brand” and (1 = not at all – 7 = definitely). Measuring justice in this manner provides robust test of our hypothesis as it allows us to demonstrate asymmetric effects of justice as a function of the alert valence. Thus, a positive score on this scale indicated a desire to punish while a negative score indicated a desire to reward the brand. In addition to the measured variables, we also computed several variables based on the full-length reviews participants chose to read. We computed the *average valence of the reviews read* by participants, the *number of recommended reviews read*, and whether participants chose to *read the not recommended review*.

## Results

*Brand Rating.* H1 posits that relative to when a fake review alert is absent, a fake negative (positive) review alert increases (decreases) brand ratings (Figure 2 shows these findings). A 2 (fake review alert: negative, positive) x 3 (time period: before alert, during alert, after alert) repeated measures mixed ANOVA on brand rating yielded a significant main effect of the alert ( $F(1, 276) = 19.99; p < .001$ ), qualified by a significant interaction ( $F(1, 276) = 43.96; p < .001$ ). The main effect of the time period was not significant ( $p > .10$ ). Planned contrasts were conducted within each alert condition across time periods. Consistent with H1, for brands which received a fake positive review alert, brand ratings decreased from *before* to *during* the active alert ( $M_{\text{before}} = 2.62, M_{\text{during}} = 2.44; F(1, 133) = 7.22; p = .008$ ). This suggests that participants punished brands which received fake positive reviews. Consistent with Study 1, *after* a fake positive review alert was removed, the brand ratings returned to similar levels as they were before the alert ( $M_{\text{after}} = 2.64; F(1, 134) = 9.33; p = .003$ ), suggesting that the effect of a fake positive review alert did not persist once the alert was removed, as ratings *before* and *after* the alert were not significantly different ( $F(1, 133) = .122; p > .70$ ).

Consistent with H1, for brands which received a fake negative review alert, brand ratings increased from *before* to *during* the active alert ( $M_{\text{before}} = 2.63, M_{\text{during}} = 3.02; F(1, 143) = 47.34; p < .001$ ). This suggests that participants discounted negative information about the brand. *After* a fake negative review alert was removed, the brand ratings decreased ( $M_{\text{after}} = 2.82; F(1, 143) = 12.77; p < .001$ ) but remained higher than *before* the alert ( $F(1, 143) = 8.71; p = .004$ ), demonstrating a persistent effect of a fake negative review alert. It is possible that the fake negative review alert was more memorable to



participants or that it led them to attempt retroactive correction for the brand rating they assigned in the first time period. While we were agnostic as to the longitudinal effects that emerged in round 3, this finding emerged as a potentially interesting avenue for future research.

-----

Insert figure 10 about here

-----

*Perception of Average Rating Accuracy.* H2 posits that the presence of a fake review alert, regardless of valence, decreases the perception of average rating accuracy. A 2 x 3 repeated measures mixed ANOVA on the perception of average rating accuracy yielded a significant main effect of time period ( $F(1, 276) = 6.20; p = .013$ ), qualified by a significant interaction ( $F(1, 276) = 6.20; p = .013$ ). The main effect of the alert was not significant ( $p > .15$ ). Planned contrasts were conducted within each alert type. Surprisingly, perceptions of the average rating accuracy was not affected by the fake negative review alert ( $p > .15$ ). However, as expected, within the fake positive review alert condition, the average ratings were perceived to be more accurate before versus during the active alert ( $M_{\text{before}} = 4.63, M_{\text{during}} = 4.05; F(1, 132) = 16.89; p < .001$ ). After the alert was removed, the perceived accuracy of the average rating marginally increased ( $M_{\text{after}} = 4.29; F(1, 132) = 3.62; p = .059$ ), but this was still significantly lower than before the alert ( $F(1, 132) = 6.91; p = .01$ ). Taken together, these findings suggest that consumers attend more to accuracy perceptions in the presence of a fake positive versus

negative review alert, and also suggest that fake positive review alerts have persistent effects, on both brand ratings and the perception of ratings accuracy, even once they are removed.

-----

Insert figure 11 about here

-----

*Desire for Justice.* H3 posits that a fake positive review alert increases the desire for justice relative to other conditions. A 2 x 3 repeated measures mixed ANOVA on the desire for justice yielded significant main effects of the alert ( $F(1, 276) = 14.52; p < .001$ ) and time period ( $F(1, 276) = 12.81; p < .001$ ), qualified by a significant interaction ( $F(1, 276) = 41.49; p < .001$ ). Planned contrasts were conducted within each alert type. In the fake negative review alert condition, the desire for justice was significantly higher before than during the active alert ( $M_{\text{before}} = .54, M_{\text{during}} = -.81; F(1, 143) = 40.95; p < .001$ ), indicating a greater desire to reward the brand in the presence of the fake negative review alert. This may suggest that consumers attempted to retroactively correct for the ratings they assigned in Round 1 when they learn that the brand has been a victim of fake negative reviews, possibly because they think they may have been influenced by fake reviews in the previous round. After the alert was removed, this effect was attenuated ( $M_{\text{after}} = .15; F(1, 143) = 43.08; p < .001$ ), but remained less negative than before the alert being active ( $F(1, 143) = 4.49; p = .036$ ). Consistent with H3, in the fake positive review alert condition, the desire for justice was significantly higher during than before the

active alert ( $M_{\text{before}} = .37$ ,  $M_{\text{during}} = -.83$ ;  $F(1, 133) = 43.08$ ;  $p < .001$ ), indicating a desire to punish the brand when the alert is active. After the alert was removed, this desire for justice did not significantly decrease ( $M_{\text{after}} = .63$ ;  $F(1, 133) = 1.06$ ;  $p > .30$ ), however there was also no difference in the desire for justice before and after alert was active ( $p > .10$ ). Taken together, these results suggest that justice may lead to increased brand evaluations, when consumers perceive that their prior brand perceptions were biased by fake reviews.

-----

Insert figure 12 about here

-----

*Average Valence of Reviews Read.* H5 posits that the average valence of reviews read will decrease in the presence of a fake positive versus negative review alert. A 2 x 3 repeated measures mixed ANOVA on the average valence of reviews read yielded significant main effects of the alert ( $F(1, 269) = 11.72$ ;  $p = .001$ ) and time period ( $F(1, 269) = 4.88$ ;  $p = .028$ ), qualified by a marginal interaction ( $F(1, 269) = 2.80$ ;  $p = .095$ ). Planned contrasts within each alert type were conducted. In the fake negative review alert condition, relative to before ( $M_{\text{before}} = 2.63$ ), participants read significantly more positive reviews during ( $M_{\text{during}} = 2.94$ ;  $F(1, 139) = 17.43$ ;  $p < .001$ ) and after ( $M_{\text{after}} = 2.84$ ;  $F(1, 139) = 17.43$ ;  $p < .001$ ) the active alert. There was no difference during and after the active alert ( $p > .20$ ). In the fake positive review alert condition, there was no effect of the alert on review reading behavior ( $p > .70$ ). Thus, this study demonstrated that, in a

longitudinal context, fake negative review alerts increase the average valence of reviews read relative to fake positive review alerts. This could be because most participants seemed to have an overall negative impression of the restaurant, thus their baseline reading behavior established a floor, or that consumers are likely to engage in retroactive correction when they encounter a fake negative review alert, but not a fake positive review alert.

-----

Insert figure 13 about here

-----

*Number of Recommended Reviews Read.* A 2 (fake review alert: negative, positive) x 3 (time period: before alert, during alert, after alert) repeated measures mixed ANOVA on the number of recommended reviews read yielded a significant main effect of time period ( $F(1, 279) = 16.79; p < .001$ ). Neither the main effect of the alert type nor the interaction were significant ( $p > .15$ ). Participants read significantly fewer reviews after the active alert ( $M_{\text{after}} = 3.59$ ) than before ( $M_{\text{before}} = 4.36; F(1, 279) = 16.79; p < .001$ ) and during ( $M_{\text{after}} = 4.36; F(1, 279) = 31.25; p < .001$ ). There was no significant difference before and during the alert being active ( $p > .90$ ). This suggests that participants required fewer reviews to reach their decisions as the exposures to the brand increased. Interestingly, there was no difference between before and during the active alert, suggesting that the presence of an alert may increase participants' need to acquire additional information.

*Did Participants Read the “Not Recommended” Review?* The effects of the fake review alert and time period did not significantly impact participants’ likelihood of reading the not recommended review ( $p > .10$ ). On average, 34% of participants chose to read the not recommended review in at least one time period.

### *Discussion*

The findings of this study provide support for H1 by replicating the effects in Study 1 while also demonstrating the effect of fake negative review alerts. Furthermore, this study provided internal validity to bolster the findings of Study 1 by using a fictitious brand and avoiding concerns of self-selection as all participants were asked to rate the brand. However, it is important to note that this study involved the same participants across three rounds, where it is unlikely that the same consumers reviewed the brand across each time period in Study 1. As such, the repeated-measures element of this study may be able to account for some of the unpredicted effects witnessed in this study. For example, our theory posits that the desire for justice is significantly lower (indicating a desire to punish) for a fake positive review alert, relative to negative or absent. Yet, our results demonstrated that the fake negative review alert led to a significantly higher desire for justice (indicating a desire to reward), relative to positive or absent. We posit that this occurred as participants engaged in a retroactive correction process in the latter rounds to correct for Round 1. Although participants generally held an unfavorable opinion of the brand in Round 1, in Round 2 it appeared that a fake positive review alert merely reinforced this viewpoint. On the other hand, a fake negative review alert may have led participants to attribute their initial brand rating to undue influence from fake reviews, which caused them to correct for Round 1 in the latter

rounds. Although tangential to the focus of this paper, this would suggest that consumers may more strongly attempt to correct past actions when they were effectively manipulated into discounting a brand. This would also inform the persistent effects of a fake negative review alert even after it is removed for consumers who were impacted by it in prior periods (e.g., lingering effects of a guilty conscience). While it is unlikely that consumers would review a brand in multiple periods, this is an area for future research which may have implications for repeat consumers.

In the remaining studies, we will focus on a single time period to assess the effect of fake review alerts between-subjects. In doing so, we can isolate the effect of a fake review alert on consumers from their previous brand familiarity and brand dispositions.

### **Study 3: The Dual Mediating Roles of Perception of Average Rating Accuracy and Desire for Justice**

Study 3 provides a direct test of H4 in that we demonstrate the mediation paths of the perception of average rating accuracy and the desire for justice on fake review alert effects on brand ratings. While the previous study demonstrated the effects of fake review alerts on these variables independently, due to the complexity of longitudinal mediation analysis, we did not directly test this proposition. In addition to replicating the findings of prior studies, and demonstrating the mediating roles of both average rating accuracy and the desire for justice, in this study we seek to further expand the robustness of our findings by using a different category as well, restaurants.

#### *Procedure*

One-hundred and fifty participants (68% female;  $M_{\text{age}} = 21.15$ ; undergraduate negative, positive) between-subjects design. The sample size was a convenience sample

based on the undergraduate participants for a one-week time period. Participants read a scenario in which they were looking for a new café and found one that seemed suitable. Participants viewed summary information of the restaurant detailing the price range (\$), average rating (3.5 out of 5 stars), number of reviews (85), and a small picture of the café interior. They were told that they could view some reviews for the brand before making a decision. As before, participants had access to 11 reviews (10 “recommended” and one “not recommended”) and could read as few or as many reviews as they liked. After reading their reviews, participants responded to the same measures as Study 2. See Appendix C for stimuli.

### *Results*

*Brand Rating.* A 3-cell (fake review alert: absent, negative, positive) ANOVA on brand ratings yielded a significant effect ( $F(2, 147) = 35.30; p < .001$ ). Planned contrasts demonstrated that relative to when an alert was absent ( $M_{\text{absent}} = 3.10$ ), a fake negative review alert increased brand ratings ( $M_{\text{negative}} = 3.43; t(147) = 3.11; p = .002$ ) while a fake positive review alert decreased brand ratings ( $M_{\text{positive}} = 2.55; t(147) = -5.20; p < .001$ ). This provided support for H1.

-----

Insert figure 14 about here

-----

*Perception of Average Rating Accuracy.* H2 posits that a fake review alert, regardless of valence, decreases the perception of average rating accuracy. A 3-cell (fake

review alert: absent, negative, positive) ANOVA on perception of average rating accuracy yielded a significant effect ( $F(2, 147) = 20.65; p < .001$ ). Planned contrasts demonstrated that a fake negative review alert decreased the perception of average rating accuracy relative to absent condition ( $M_{\text{absent}} = 5.12; M_{\text{negative}} = 4.24; t(147) = -3.95; p < .001$ ). Furthermore, a fake positive review alert decreased perceptions of accuracy even further than that of the negative alert ( $M_{\text{positive}} = 3.70; t(147) = -2.42; p = .017$ ). Thus, both fake review alerts decreased the perception of average rating accuracy relative to when the alert was absent, supporting H2..

-----

Insert figure 15 about here

-----

*Desire for Justice.* H3 posits that the desire for justice increases in the presence of fake positive review alert relative to when it is negative or absent. A 3-cell (fake review alert: absent, negative, positive) ANOVA on brand ratings yielded a significant effect ( $F(2, 145) = 16.49; p < .001$ ). Planned contrasts demonstrated an asymmetric effect of the alert on participants' desire for justice such that there was no difference between when the alert was absent or negative ( $p > .25$ ). However, the desire for justice was significantly greater in the presence of a fake positive review alert ( $M_{\text{positive}} = .60$ ) relative to when the alert was absent ( $M_{\text{absent}} = -.84; t(145) = 4.29; p < .001$ ) or a fake negative alert ( $M_{\text{negative}} = -1.23; t(145) = 5.44; p < .001$ ), supporting H3. This demonstrated that



participants' desire to punish the brand was greater in presence of a fake positive review alert, relative to when the alert was absent or negative.

-----

Insert figure 16 about here

-----

*Mediation via Perception of Average Rating Accuracy and Desire for Justice.* H4 proposes a dual mediation process such that consumers' perception of the accuracy of the average rating and their desire for justice mediate the effect of fake review alerts on brand ratings as a function of the valence of the fake reviews. We tested this proposition using the PROCESS macro (model 4: Preacher, Rucker, and Hayes 2007) using the alert type (absent, negative, positive) as our independent variable, brand rating as our dependent variable, and average rating accuracy and desire for justice as our mediators. Results demonstrated significant omnibus mediation effects through both average rating accuracy ( $\beta = .02$ ;  $CI_{95\%} = [.003, .042]$ ) and the desire for justice ( $\beta = -.03$ ;  $CI_{95\%} = [-.048, -.011]$ ). Dummy coding our independent variable demonstrated the effects of both alert types relative to when an alert was absent. When consumers encountered a fake negative review alert, only mediation through average rating accuracy remained significant (average rating accuracy ( $\beta = -.08$ ;  $CI_{95\%} = [-.166, -.017]$ ), desire for justice ( $\beta = .06$ ;  $CI_{95\%} = [-.032, .179]$ ). Yet, when consumers encountered a fake positive review alert, both pathways were significant (average rating accuracy ( $\beta = -.12$ ;  $CI_{95\%} = [-.252, -.019]$ ), desire for justice ( $\beta = -.22$ ;  $CI_{95\%} = [-.380, -.117]$ ). This finding provides evidence

for H4 as accuracy perceptions always mediate but the desire for justice only mediates when consumers encounter fake positive review alerts, where they can attribute the manipulative intent to a specific brand.

*Average Valence of Reviews Read.* A 3 (fake review alert: absent, negative, positive) ANOVA on average valence of reviews read yielded a significant effect ( $F(2, 147) = 3.11; p = .048$ ). Planned contrasts further demonstrated that relative to when an alert was absent ( $M_{\text{absent}} = 2.82$ ), a fake positive review alert decreased the valence of reviews read ( $M_{\text{positive}} = 2.38; t(147) = -2.38; p = .019$ ) while a fake negative review alert did not ( $M_{\text{negative}} = 2.72; t(147) = -1.83; p > .55$ ), supporting H5. Furthermore, we tested this as a mediating variable and demonstrated that it did not mediate (i.e., the 95% confidence interval included zero), ruling out a possible alternative explanation in that participants' brand ratings might be a direct result of averaging the reviews they have read.

*Number of Recommended Reviews Read.* Consistent with Study 2, the effect of the fake review alert did not significantly impact participants' likelihood of reading the not recommended review ( $p > .95$ ). On average, participants read 3.39 reviews.

*Did Participants Read the "Not Recommended" Review?* Consistent with Study 2, the effect of the fake review alert did not significantly impact participants' likelihood of reading the not recommended review ( $p > .60$ ). On average, 20% of participants chose to read the not recommended review.

-----

Insert figure 17 about here

## *Discussion*

This study provided support for H4, by demonstrating the effect of a fake review alert on brand ratings was mediated by the perception of average rating accuracy (regardless of valence), and the desire for justice (for fake positive review alerts).

Participants are always likely to correct for any perceived inaccuracies in the average ratings, but the desire for justice yielded additional influence in the presence of positive fake review alerts relative to the other conditions. We argue that this occurs because consumers can attribute the fake reviews to the brand in the fake positive review alert condition, whereas in the fake negative review alert condition, the brand is not at fault.

We also provided evidence additional evidence for several of our hypotheses. Participants increased brand ratings in the presence of a fake negative review alert and decreased brand ratings in the presence of a fake positive review alert (H1). We demonstrated that this was mediated by both participants' perceptions of the average rating accuracy and their desire for justice. Furthermore, we demonstrated that participants read significantly more negative information in the presence of a fake positive review alert relative to other conditions (H5), but their brand ratings were not mediated by the valences read, ruling out a possible alternative explanation.

### **Study 4: The Moderating Effect of Salient Aggregate Information**

Thus far, we have demonstrated that fake review alerts influence brand ratings by via perceptions of average rating accuracy and the desire for justice. If our process is correct, the effect of a fake review alert on brand ratings can be attenuated by

emphasizing the accuracy of the remaining reviews. In doing so, the need to correct is attenuated, as well as the desire for justice, and as a result, so is the effect of the fake review alert. Study 4 provides a test of this proposition (H6), by reinforcing the accuracy of the brand rating and review volume after the fake review alert is encountered.

### *Design and Procedure*

Three-hundred and ten participants (50% female;  $M_{age} = 38.27$ ; Amazon mTurk sample; \$0.50 payment) were randomly assigned to a cell in a 2 (fake review alert: negative, positive) x 2 (aggregate information disclaimer: absent, present) between-subjects design. The sample size was determined a priori based on a 75-subject rule-of-thumb for online samples at the time the study was conducted. Having already established the directional effects of the fake review alert valences, in this study we only used the negative and positive conditions. Participants viewed the same stimuli for the hotel used in Study 2. To manipulate the salience of the aggregate information, we placed a disclaimer at the bottom of the alert which read “We have removed the fraudulent reviews, resulting in a 3.2 out of 5.0 rating, based on 264 reviews” in the present aggregate information disclaimer condition. Participants followed the same procedure as in the previous study and the same measures were assessed. See Appendix C for the stimuli.

### *Results*

*Brand Ratings.* H6 posits that presenting a brand’s aggregate information after a fake review alert is displayed attenuates the effect of the fake review alert on brand ratings. A 2 (fake review alert: negative, positive) x 2 (aggregate information disclaimer: absent, present) ANOVA on brand ratings yielded a significant main effect of the alert

( $F(1, 306) = 48.27; p < .001$ ), qualified by a significant interaction ( $F(1, 306) = 12.69; p < .001$ ). The main effect of the aggregate information disclaimer was not significant ( $p > .80$ ). In the absence of the aggregate information disclaimer, a fake positive review alert ( $M_{\text{positive}} = 2.08$ ) decreased brand ratings relative to a fake negative review alert ( $M_{\text{negative}} = 3.03; F(1, 306) = 55.28; p < .001$ ). This replicated our prior findings. In the presence of the aggregate information disclaimer, the effects of the alerts were attenuated ( $M_{\text{positive}} = 2.43, M_{\text{negative}} = 2.73; F(1, 306) = 5.72; p = .017$ ). This finding provides support for H6 as it demonstrates that the effects of fake review alerts are attenuated in the presence of salient aggregate information.

-----

Insert figure 18 about here

-----

*Perception of Average Rating Accuracy.* A 2 (fake review alert: negative, positive) x 2 (aggregate information disclaimer: absent, present) ANOVA on the perception of average rating accuracy yielded significant main effects of the alert ( $F(1, 306) = 14.97; p < .001$ ) and the aggregate information disclaimer ( $F(1, 306) = 5.45; p = .02$ ), qualified by a significant interaction ( $F(1, 306) = 5.41; p = .021$ ). In the absence of the aggregate information disclaimer, the perception of average rating accuracy was significantly lower in the positive versus negative alert condition ( $M_{\text{negative}} = 4.70; M_{\text{positive}} = 3.62; F(1, 306) = 19.21; p < .001$ ). Consistent with prior studies, participants felt that the average rating was less accurate in the presence of a positive versus negative

alert. Conversely, in the presence of the aggregate information disclaimer, the alerts had no significant effect on the perception of average rating accuracy ( $M_{\text{negative}} = 4.70$ ,  $M_{\text{positive}} = 4.43$ ;  $F(1, 306) = 1.19$ ;  $p > .25$ ), as expected. This suggests that by re-emphasizing the average rating after participants viewed an alert, the effect of the alert on the perception average rating accuracy was attenuated.

-----

Insert figure 19 about here

-----

*Desire for Justice.* A 2 (fake review alert: negative, positive) x 2 (aggregate information disclaimer: absent, present) ANOVA on desire for justice yielded a significant main effect of the alert ( $F(1, 306) = 30.86$ ;  $p < .001$ ), qualified by a significant interaction ( $F(1, 306) = 11.31$ ;  $p = .001$ ). The main effect of aggregate information disclaimer was not significant ( $p > .10$ ). In the absence of the aggregate information disclaimer, the desire for justice was significantly greater in the presence of a fake positive versus negative review alert ( $M_{\text{negative}} = -.88$ ;  $M_{\text{positive}} = 1.28$ ;  $F(1, 306) = 39.81$ ;  $p < .001$ ), consistent with prior studies. Conversely, in the presence of the aggregate information disclaimer, the alerts had no significant effect on the desire for justice ( $M_{\text{negative}} = -.45$ ,  $M_{\text{positive}} = .09$ ;  $F(1, 306) = 2.40$ ;  $p > .10$ ). This suggests that the aggregate information disclaimer refocuses participants' attention on the aggregate information, attenuating their desire for justice.

-----  
Insert figure 20 about here  
-----

*Moderated-Mediation of Brand Ratings via the Perception of Average Rating Accuracy and Desire for Justice.* A PROCESS Model 8 analysis (Preacher, Rucker, and Hayes 2007) with the fake review alert and the aggregate information disclaimer as independent variables, brand rating as the dependent variable, and the perception of average rating accuracy and desire for justice as the mediators, yielded significant indices of moderated-mediation (perception average rating accuracy:  $\beta = .14$ ;  $CI_{95\%} = [.026, .307]$ ; desire for justice:  $\beta = .21$ ;  $CI_{95\%} = [.087, .382]$ ).

In the absence of the aggregate information disclaimer, both the perception of average rating accuracy ( $\beta = -.18$ ;  $CI_{95\%} = [-.329, -.095]$ ) and desire for justice ( $\beta = -.28$ ;  $CI_{95\%} = [-.428, -.159]$ ) mediated the effect of an alert on consumers' brand ratings. As expected, participants used both their perception of average rating accuracy and their desire for justice to inform their brand ratings when positive fake review alerts were present. Yet, in the presence of the aggregate information disclaimer, both pathways were no longer significant (perception of average rating accuracy:  $\beta = -.05$ ;  $CI_{95\%} = [-.141, .029]$ ; desire for justice:  $\beta = -.07$ ;  $CI_{95\%} = [-.178, .016]$ ). This suggests that by making the aggregate information salient after consumers view an alert, a website could ease concerns about rating accuracy and also attenuate their consumers' desire for justice. These results provide additional evidence for H5, as we demonstrated the mediation process once more, and how it can be attenuated.

*Average Valence of Reviews Read.* A 2 (fake review alert: negative, positive) x 2 (aggregate information disclaimer: absent, present) ANOVA on the average valence of reviews read yielded a significant effect of the alert ( $F(1, 306) = 10.40; p = .001$ ). Neither the main effect of the aggregate information disclaimer, nor the interaction term, were significant ( $p > .30$ ). On average, participants read more negative reviews when they encountered a fake positive review alert ( $M_{\text{positive}} = 2.54$ ) relative to when they viewed a fake negative review alert ( $M_{\text{negative}} = 2.89$ ). In testing for the mediating role of this variable, we once again rule out this alternative explanation (i.e., the confidence interval included zero).

*Number of Recommended Reviews Read.* A 2 (fake review alert: negative, positive) x 2 (aggregate information disclaimer: absent, present) ANOVA on the number of recommended reviews read yielded a significant main effect of the alert ( $F(1, 306) = 6.95; p = .009$ ). Neither the main effect of the aggregate information disclaimer, nor the interaction term, were significant ( $p > .20$ ). On average, participants read more reviews when they encountered a fake negative review alert ( $M_{\text{negative}} = 3.90$ ) relative to when they viewed a fake positive review alert ( $M_{\text{positive}} = 3.28$ ). While not a replication of our previous finding, this may suggest category effects (e.g., hotel versus restaurant) when reading reviews, though this is left as an interesting avenue for future research.

*Did Participants Read the “Not Recommended” Review?* A 2 (fake review alert: negative, positive) x 2 (aggregate information disclaimer: absent, present) binary logistic regression on the likelihood of reading the not recommended review yielded a significant main effect of the alert ( $\chi^2(1) = 8.60; p = .003$ ). Neither the main effect of the aggregate information disclaimer, nor the interaction term, were significant ( $p > .70$ ). Participants



were more likely to read the not recommended review when they viewed a fake negative review alert ( $P_{\text{negative}} = .31$ ) versus a fake positive review alert ( $P_{\text{positive}} = .11$ ), suggesting that participants were more curious about fake negative, versus positive, reviews. Once again, while this did not replicate the findings of our previous studies, it does provide the basis for future exploration of category effects on review reading behavior.

### *Discussion*

This study provided support for H6 by demonstrating that the negative effects of fake review alerts (i.e., leading consumers' expectations to deviate from the average ratings) could be attenuated by reinforcing the aggregate information after consumers encounter the alert. This holds managerial relevance for websites that choose to use fake review alerts, as the alerts may be biasing their consumers' judgments. Furthermore, this study provided additional evidence for the mediating roles of the perception of average rating accuracy and the desire for justice by demonstrating significant moderated-mediation of their influences on brand ratings. By refocusing consumers' attention on the aggregate statistics, they are less likely to perceive the average ratings as inaccurate and also less likely to seek justice.

One interesting finding of this study stems from the differing reading behaviors of the reviews across studies. While prior studies largely showed null effects of fake review alerts on the number of reviews read and whether participants read the not recommended review, this study did not. This is perhaps due to the different study paradigms (i.e., longitudinal versus static) or categories (e.g., hotels versus restaurants) or perhaps a combination of the two. In this study, participants were significantly more likely to read the not recommended review when it was a negative, versus positive, review. This could

be due to the perceived relative novelty of this review type versus fake positive reviews, or it could be because consumers think they can extract diagnostic information from these reviews. While outside the scope of this paper, the factors which influence the choice to read specific reviews is a rather interesting future direction.

## **GENERAL DISCUSSION**

Websites which host reviews are attempting to improve their consumers' experiences by removing fake reviews. In doing so, they must expertly navigate perceptions of fraudulence and censorship. To do so, some websites, like Yelp, use fake review alerts to increase the transparency with their consumer base. However, our findings demonstrate that this attempt at transparency may backfire by biasing consumers' decisions. That is, these alerts cause consumers' opinions of the brands to deviate significantly from the honest reviews which have evaluated the brand.

Study 1 provided empirical support for the effect of fake positive review alerts on brand ratings. Study 2 replicated these findings in the lab while also demonstrating the effects of a fake negative review alert. Study 3 demonstrated the dual mediation process via the perception of average rating accuracy and the desire for justice on brand ratings, while also demonstrating the effect of an alert on reading behaviors. Study 4 then demonstrated one such way that managers can attenuate the bias created by fake review alerts; by refocusing consumers' attention on the average brand rating and review volume after encountering a fake review alert.

### **Theoretical Implications**

Prior research has largely demonstrated that the activation of persuasion knowledge decreases brand attitudes when the marketing tactics are salient (Campbell 1995; Campbell and Kirmani 2000) or increases brand attitudes when the marketing tactic is credible (Isaac and Grayson 2017). In this work, we demonstrate that consumers not only hold accuracy goals in the presence of fraudulent information, but also justice goals against the source of the manipulative behavior. Rather than merely discount the fraudulent information, we demonstrate that consumers desire justice against those responsible and will actually punish or reward a brand depending on its role in the fake reviews. Unfortunately, this creates a bias as consumers overcorrect for the fake review alerts, perhaps increasing or decreasing a consumer's expectation beyond the actual experience they will receive. However, we demonstrate that by refocusing consumers' attention on the aggregate brand information after encountering an alert, consumers' brand ratings are more in line with the average.

### *Managerial Implications*

As consumers' demand for veracity increases, websites will have to adapt how, and what, information is provided to consumers. In doing so, the information websites provide will have the ability to significantly influence how consumers perceive brands, but also how consumers perceive the websites themselves. Regarding brand attitudes, websites must carefully monitor the reviews listed so as to provide consumers accurate information about the brand. Fake reviews have the ability to yield expectation disconfirmation which may long-term consequences for brand loyalty, but also website loyalty if consumers find the information to be inaccurate. Thus, websites should be focused on providing consumers accurate information rather than guiding perceptions of

expected accuracy, because once consumers actually use a product they will be able to contrast their experience with the expectation derived from the information the website provided.

### Limitations and Future Research

This paper provided both empirical data from web-scraped reviews and experiments to provide a robust account of the effects of fake review alerts on consumers. However, as with any non-controlled data source, it is impossible to control for consumers' previous exposures to the brands in the web-scraped data. Thus, the persons visiting the brands and writing the reviews self-select into these decisions and we cannot account for that aspect. Future research would benefit from a field experiment that is able to exogenously determine when the alerts are active, and for whom. Furthermore, because fake review alerts are a relatively new phenomena, their long-term effects on businesses are not apparent. It would be interesting to study brand-side responses to these alerts to determine if brands attempt to improve product quality, or use other marketing variables, to overcome the publicity from an alert.

From an experimental viewpoint, while these studies allude to competitive environments, they do not provide additional information about competitors. In the real world, consumers have access to reviews from several brands to shape their decisions. It would thus be interesting to understand the effects of one brand's fake review alerts on the evaluations of competing brands. Would consumers discount the ratings of all other brands when the focal brand has received a fake negative review alert, and would they elevate the ratings of all other brands when the focal brand has received a fake positive

review alert? This is perhaps an empirical question that could be tested in conjunction with future research of how brands should respond to fake review alerts.

Another interesting finding that arose from this research is that the participants consistently assigned brand ratings beneath that of the average rating in the control conditions. In unpublished work from one of the authors, they examine a review sequence bias on expected brand ratings, and the authors believe that this provides further evidence. For example, even if a product has an average rating of 4.0, if the first review a consumer encounters is a 1.0, their brand rating will likely be lower than that of the average. Thus, exploring the weighting of this effect relative to fake review alerts could be interesting.

More broadly, reviews are used by consumers as sources of information to help them form decisions. How would alerts impact the effectiveness of other information sources (e.g., news articles or social media posts)? Consumers' pre-held beliefs to both news sources and social media creators may have interesting interactions with fake information alerts. Furthermore, given the topical importance of fake news claims, this avenue provides an important future direction for this research.

## APPENDIX A – Essay I Stimuli

### Study 1 Low Volume Stimuli (other volumes listed in Table 1)

#### BUYING HEADPHONES ONLINE

Listening to music makes you happy and spirited throughout the day. You want to purchase a new set of headphones that you can wear for an extended period of time each day. You decide to do some comparison shopping online, and so you visit [www.amazon.com](http://www.amazon.com).

After some research, you have narrowed down your choices to two different sets of headphones that meet your needs. They each have noise-cancelling functionality, over-the-ear-padding, and an extended cord. You plan on buying one of these soon. Please indicate your level of preference:

A



**Audio-Technica Ath-M30. \$58.98.**  
**Review Score: 3.5/5.0.**  
**Number of Reviews: 8.**

Very clean, deep bass response

Great for project studio engineers  
and recording musicians

—  
STRONGLY  
PREFER A



PREFER A



SLIGHTLY  
PREFER A



NO  
PREFERENCE



SLIGHTLY  
PREFER B



PREFER B



**Sennheiser HD 202. \$58.69.**  
**Review Score: 3.3/5.0.**  
**Number of Reviews: 84.**

Professional-quality dynamic  
stereo headphones

Ideal for DJ's and audio pros

—  
STRONGLY  
PREFER B



#### BUYING A COFFEE MAKER ONLINE

Consuming coffee will keep you focused and help you perform better in school. You want to purchase a coffee maker to make coffee at home each morning. You decide to do some comparison shopping online, and so you visit [www.amazon.com](http://www.amazon.com).

After some research, you have narrowed down your choices to two different machines that meet your needs. They each have 12-Cup capacity, timed start, warming capabilities, and a digital display. You plan on buying one of these machines soon. Please indicate your level of preference:

A



**Cuisinart DCC 1200.**  
**\$47.89.**  
**Review Score: 3.4/5.0.**  
**Number of Reviews: 6.**  
Classic brushed metal  
design

Ergonomic handle for  
comfortable, dripless  
pouring

—  
STRONGLY  
PREFER A



PREFER A



SLIGHTLY  
PREFER A



NO  
PREFERENCE



SLIGHTLY  
PREFER B



PREFER B



B



**Hamilton Beach 46201.**  
**\$48.89.**  
**Review Score: 3.0/5.0.**  
**Number of Reviews: 68.**  
Easy-access design for fast  
filling

Swivel base for fast, easy  
access to the water reservoir

—  
STRONGLY  
PREFER B



### BUYING A MICROWAVE ONLINE

You generally study best after you eat food. You want to purchase a microwave so that you can quickly prepare hot meals at home. You decide to do some comparison shopping online and so you visit [www.amazon.com](http://www.amazon.com).

After some research, you have narrowed down your choices to two different microwaves that meet your needs. They each have a defrost setting, a rotating plate, and a digital clock. You plan on buying one of these microwaves soon. Please indicate your level of preference:

A



**RCA Microwave Oven.**  
\$45.49.  
Review Score: 3.4/5.0.  
Number of Reviews: 9.  
8 one-touch functions

Cook and defrost by weight

—  
STRONGLY  
PREFER A  
☐

PREFER A  
☐

SIGHTLY  
PREFER A  
☐

NO  
PREFERENCE  
☐

SIGHTLY  
PREFER B  
☐

PREFER B  
☐

B



**Hamilton Beach Microwave  
Oven.** \$46.39.  
Review Score: 3.1/5.0.  
Number of Reviews: 62.  
Chrome finish

Multi-function settings

—  
STRONGLY  
PREFER B  
☐

### BUYING A SPEAKER SYSTEM ONLINE

You love listening to music when at home, so you want to purchase a small speaker system for your room. You decide to do some comparison shopping online, and so you visit [www.amazon.com](http://www.amazon.com).

After some research, you have narrowed down your choices to two different speaker systems that meet your needs. They each have extended cords, a sub-woofer, and USB connectivity. You plan on buying one of these speaker systems soon. Please indicate your level of preference:

A



**Logitech S150.** \$38.69.  
Review Score: 3.5/5.0.  
Number of Reviews: 12.

Detachable units

Wide bases  
—  
STRONGLY  
PREFER A  
☐

PREFER A  
☐

SIGHTLY  
PREFER A  
☐

NO  
PREFERENCE  
☐

SIGHTLY  
PREFER B  
☐

PREFER B  
☐

B



**SoundSoul AL-101.** \$39.29.  
Review Score: 3.2/5.0.  
Number of Reviews: 74.

Swivel Base

Built-In Controls  
—  
STRONGLY  
PREFER B  
☐

## BUYING A LOUNGE CHAIR ONLINE

You love relaxing by watching TV in a comfortable chair. You want to purchase a new lounge chair to use in your home. You decide to do some comparison shopping online, and so you visit [www.amazon.com](http://www.amazon.com)

After some research, you have narrowed down your choices to two different chairs. They each have firm arm rests, head rests, and ergonomic back support. You plan on buying one of these chairs soon. Please indicate your level of preference:

A



**Herman Miller Ergo.**  
**\$48.98.**  
**Review Score: 3.4/5.0.**  
**Number of Reviews: 5**

Easy Recline

Plush cushioning

---  
STRONGLY  
PREFER A



PREFER A



SIGHTLY  
PREFER A



NO  
PREFERENCE



SIGHTLY  
PREFER B



PREFER B



B



**Eurotech Manchester.**  
**\$69.98**  
**Review Score: 3.1/5.0.**  
**Number of Reviews: 98.**

One-touch recliner

Textured arm rests

---  
STRONGLY  
PREFER B





## Study 2 – Low Volume Stimuli (other volumes listed in table 1)



Which option would you prefer?

A	B	C	D
---	---	---	---

Would you be more likely to purchase one of the options pictured here, or defer purchase and look elsewhere?

Purchase one of these options

Defer Purchase, and look elsewhere

How would you classify the amount of information provided?

Not enough information						Too much information
------------------------	--	--	--	--	--	----------------------

Study 3 – Low Volume Stimuli (other ratings and volumes listed in Table 1)

SHOP ALL DEPARTMENTS -

SEARCH:


⌵

CART (0)

[Home](#) > [Compare Products](#)

Compare Products


Remove [x]



A

Add to Cart

Remove [x]



B

Add to Cart

Name	KitchenAid® 10-Speed Reversing Motor Blender	Oster® 10-Speed Die Cast Blender
Price	\$41.99	\$41.67
Rating	3.4 / 5.0	3.1 / 5.0
# of Reviews	8	64

[measure same as Study 1]

Study 4 – Low Volume Stimuli (other ratings, volumes and differences listed in Table 1)

[SHOP ALL DEPARTMENTS](#)

SEARCH:

CART (0)

[Home](#) > [Compare Products](#)

Compare Products

A

Add to Cart

Remove [x]

B

Add to Cart

Name	Sony QuietComfort in-ear Headphones	Panasonic EcoSound in-ear Headphones
Price	\$99.98	\$99.97
Rating	3.8 / 5.0	3.0 / 5.0
# of Reviews	9	57

## Study 5 – Low Volume Stimuli (other valences, volumes listed in Table 1)

Imagine being in the market for a new hand mixer (electric blender). After some searching, you've found a few comparable options online. On the next page, please take a few moments to evaluate both options before answering the questions.

**OPTION A**



**OPTION B**



Brand	Cuisinart Hand Mixer	KitchenAid Hand Mixer
Price	\$39.97	\$39.98
Average Product Rating	1.3 / 5.0	1.0 / 5.0
Number of Reviews	13	34

Strongly Prefer  
Option A

☐

Prefer Option A

☐

Slightly Prefer  
Option A

☐

No Preference

☐

Slightly Prefer  
Option B

☐

Prefer Option B

☐

Strongly Prefer  
Option B

☐

Would you be more likely to:

- ☐ Purchase Option A
- ☐ Purchase Option B
- ☐ Defer Choice, and look for other options

## Study 6 – Low Volume Stimuli (other volumes listed in Table 1)

**Please imagine the following scenario:** In need of a new blender, you go online to do some comparison shopping. After narrowing down your choices online, you must now decide which one you would prefer to purchase.

[SHOP ALL DEPARTMENTS +](#)

SEARCH:

CART (0)

[Home](#) > [Compare Products](#)

### Compare Products [Clear all](#)

Compare Products

Remove [x]

A

Add to Cart

Remove [x]

B

Add to Cart

Name	KitchenAid® 10-Speed Reversing Motor Blender	Oster® 10-Speed Die Cast Blender
Price	\$41.99	\$41.67
Rating	3.4 / 5.0	3.0 / 5.0
# of Reviews	6	52

If you were shopping for a blender in this price range, would you:

Purchase Blender A	Purchase Blender B	Defer purchase to look for other options
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

---

How confident are you in your preference between the two blenders?

Very Unconfident						Very Confident
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

---

If you were shopping for a blender, how likely would you be to purchase one of the featured blenders?

Very Unlikely	Unlikely	Somewhat Unlikely	Undecided	Somewhat Likely	Likely	Very Likely
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please indicate the importance of each attribute in making your decision:

	Not at all Important	Very Unimportant	Somewhat Unimportant	Neither Important nor Unimportant	Somewhat Important	Very Important	Extremely Important
Image of Blender	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Product Description	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Price	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rating	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Number of Raters	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Brand	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Study 7 – Low Volumes Stimuli (other volumes listed in Table 1)

**Option A**



***RCA Microwave Oven. \$45.49.***

***Review Score: 3.4/5.0.***

***Number of Reviews: 9.***

*6 one-touch functions  
Cook and defrost by weight*

**Option B**



***Hamilton Beach Microwave Oven. \$46.39.***

***Review Score: 3.1/5.0.***

***Number of Reviews: 62.***

*Chrome finish  
Multi-function settings*

## *APPENDIX B – ESSAY I SUPPLEMENTAL STUDIES*

### *Appendix B1*

#### *Amazon Choice Set Data*

To understand just how often consumers are faced with a possible tradeoff between average product ratings and review volumes in choice sets, we analyzed a publicly-available data set from McAuley et al. (2015) which included over 142 million reviews and 2.5 million products, and respective consideration sets, for Amazon products across 24 different categories from May 1996 – July 2014. The data was split into two files, one which featured all reviews for the products, and one which featured the metadata, including the “related and also viewed” choice set options. By parsing these lists together, we were able to reconstruct the consideration sets consumers encountered. We then computed two measures from this data. First, we coded for whether any alternative choice option created a tradeoff scenario (i.e., a flip) with the focal product in which one option had a higher average product rating and fewer reviews relative to the other. We also coded for the frequency of this occurrence in each choice set. Thus, we have a measure of whether a tradeoff occurred in each set (i.e., share of choice sets flipped at least once), and how many options forced a tradeoff in that choice set (i.e., flipped share of the choice sets).

Our analysis revealed that 2,050,549 of the 2,503,422 (79%) choice sets demonstrated at least one tradeoff in the choice set between the focal product and its alternatives. Furthermore, on average, nearly half of the choice set, 8 out of 18 options (47%), had a lower rating but more reviews than the focal product.



Thus, the tradeoffs consumers face seem to be rather common in the online market. In fact, given that this information was based on Amazon data, which generally has more reviews than other online retailers, these findings are a conservative estimate of the state of the market, given the increased variance around product ratings when fewer reviews are available. Thus, for most retailers, the frequency of tradeoffs occurring is likely much greater.

Table B1 – Amazon Tradeoff Data

Category ID	Category	Number of Products	Products Flipped* at Least Once	Share of Products flipped at least once (%)	Smallest CS** size	Largest CS size	Average CS size	Smallest number of items "flipped" in CS	Largest number of items "flipped" in CS	Average number of items "flipped" in CS	Average "flipped" share***
1	Amazon Instant Video	110	83	75%	0	57	12	0	52	7	59%
2	Apps for Android	6,707	4,258	63%	0	60	17	0	58	7	39%
3	Automotive	213,414	174,078	82%	0	60	19	0	59	9	45%
4	Baby	44,517	40,179	90%	0	60	29	0	59	14	47%
5	Beauty	175,633	150,671	86%	0	60	20	0	58	10	49%
6	Books	261,207	180,081	69%	0	60	6	0	57	3	45%
7	CDs and Vinyl	87,410	60,081	69%	0	59	6	0	54	2	41%
8	Cell Phones and Accessories	72,863	57,815	79%	0	60	16	0	56	7	47%
10	Digital Music	5,832	3,325	57%	0	56	5	0	55	1	33%
11	Electronics	177,264	145,417	82%	0	60	19	0	58	9	47%
12	Grocery and Gourmet Food	109,956	93,351	85%	0	60	19	0	55	10	49%
13	Health and Personal Care	180,015	156,215	87%	0	60	24	0	57	12	49%
14	Home and Kitchen	251,964	218,096	87%	0	60	25	0	59	12	47%

15	Kindle Store	43,037	29,954	70%	0	60	11	0	52	5	41%
16	Movies and TV	34,569	24,131	70%	0	60	7	0	52	3	44%
17	Musical Instruments	32,306	25,762	80%	0	60	18	0	55	8	45%
18	Office Products	74,213	63,361	85%	0	60	24	0	59	12	47%
20	Pet Supplies	69,424	60,128	87%	0	60	24	0	57	12	48%
21	Sports and Outdoors	285,373	241,576	85%	0	60	19	0	58	9	45%
22	Tools and Home Improvement	129,065	104,454	81%	0	60	19	0	56	9	45%
23	Toys and Games	234,300	206,437	88%	0	60	31	0	60	14	46%
24	Video Games	14,243	11,096	78%	0	60	15	0	59	7	45%

\* Two products are being "flipped" if one product has higher average but lower number of ratings compared to another product.

\*\* In this context the term "Consideration Set" denotes the products identified by Amazon.com as "related and also viewed" with respect to the focal product

\*\*\* Share of the products in CS flipped with respect to the focal product

## APPENDIX B2

### *Relative versus Absolute Differences in Review Volumes*

The purpose of this study was to compare the influence of an increase in the *relative* difference between choice options on the review volume attribute (the goal of the main paper) to an increase in the *absolute* difference between choice options on the review volume attribute as demonstrated by prior work (numerosity effect; Pandelaere, Briers, and Lembregts 2011; Bagchi and Li 2010). To this end, we created the low review volumes and high review volumes conditions, where the absolute difference is held constant while relative difference changes, similar to previous studies, while also adding a high review volumes condition in which the relative difference is held constant with the low review volumes conditions, while the absolute difference changes. This last condition, therefore, provides a test of numerosity effect observed in prior work but in the context of multiple numerical attributes.

### *Method*

*Participants and design.* We recruited 153 participants ( $M_{\text{age}} = 36.07$ ; 55% female) from Amazon mTurk in exchange for a \$0.50 payment. The sample size was determined based on a 50-subject rule-of-thumb for online samples at the time the study was conducted. Participants were randomly assigned to a condition in a 3 (review volume levels: low, high relative, high absolute) between-subjects x 3 (product replicates: BBQ grills, patio furniture, patio umbrella) within-subject mixed design.

*Procedure.* We used the same procedure as in Study 1. For each product replicate, participants would indicate their relative preference between two choice options that required a tradeoff between a higher-rated, fewer reviews option and a lower-rated, more

reviews option. In addition to the low (e.g., 12 vs. 45) and high (e.g., 212 vs. 245) review volume conditions used in past studies (manipulating relative difference), we also included a high review volume (e.g., 212 vs. 795) condition (manipulating absolute difference). Following this, we also measured the need for additional information.

*Results.* A repeated measures ANOVA of review volume levels and product replicates on preference yielded main effects of both product replicates ( $F(1, 150) = 17.74; p < .001$ ) and review volume levels ( $F(1, 150) = 8.40; p < .001$ ), qualified by a significant quadratic interaction ( $F(2, 150) = 4.06; p = .019$ ). To explain this interaction, we first compare the low and high relative difference review volumes conditions, which are most similar to our prior studies.

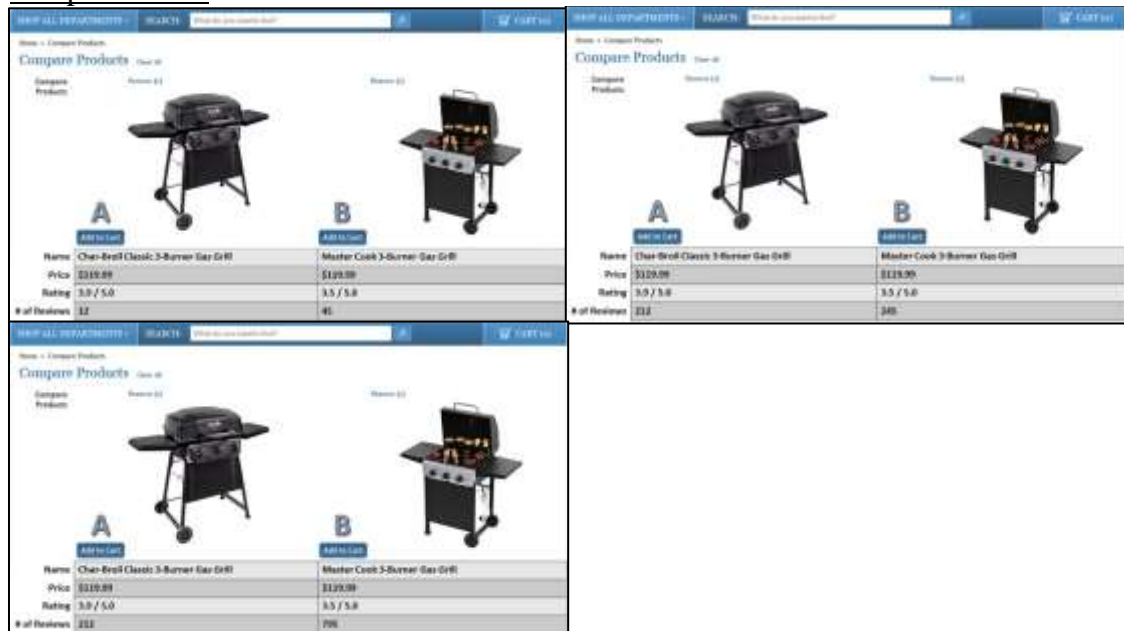
Consistent with prior results, the repeated measures ANOVA of product replicates and review volume levels yielded significant main effects of product replicates ( $F(1, 98) = 25.09; p < .001$ ), and review volume levels ( $F(1, 98) = 11.94$ ; partial  $\eta^2 = .11$ ;  $p = .001$ ), while the interaction was not significant ( $p > .65$ ). Consistent with our previous studies, preference for the higher-rated options was greater when review volumes were high ( $M_{\text{relative}} = 3.22$ ) relative to low ( $M_{\text{low}} = 4.03$ ). The main effect of product replicates merely demonstrates that consumer preference for the higher-rated, fewer reviews option is weaker for patio furniture ( $M_{\text{furniture}} = 4.43$ ) relative to the BBQ grills ( $M_{\text{grills}} = 3.11$ ;  $p < .001$ ) and the patio umbrellas ( $M_{\text{umbrellas}} = 3.33$ ;  $p < .001$ ). There was no significant difference in preference between BBQ grills and patio umbrellas ( $p > .30$ ).

By contrast, when examining low and high absolute difference review volumes levels conditions, a repeated measures ANOVA yielded a significant main effect of product replicates ( $F(1, 102) = 8.63; p = .004$ ), qualified by an interaction with product

replicates and review volume levels ( $F(1, 102) = 6.85$ ; partial  $\eta^2 = .06$ ;  $p = .01$ ). Importantly, the main effect of review volumes was not significant ( $p > .50$ ). Planned contrasts demonstrate that while BBQ grills ( $M_{\text{low}} = 3.31$ ;  $M_{\text{absolute}} = 4.02$ ;  $F(1, 102) = 4.04$ ; partial  $\eta^2 = .04$ ;  $p = .047$ ) exhibit a pattern consistent with a numerosity effect, patio furniture exhibits a reverse pattern ( $M_{\text{low}} = 4.90$ ;  $M_{\text{absolute}} = 4.23$ ;  $F(1, 102) = 2.87$ ; partial  $\eta^2 = .03$ ;  $p = .093$ ), and preference for patio umbrellas was directionally consistent with a numerosity effect ( $M_{\text{low}} = 3.86$ ;  $M_{\text{absolute}} = 4.28$ ;  $F(1, 102) = 4.04$ ; partial  $\eta^2 = .01$ ;  $p > .25$ ).

*Discussion.* The results of this study suggest that both relative and absolute differences in review volumes can affect consumer preferences in the context of integration of multiple numerical attributes, and gives initial evidence that our proposed effect (effect of relative differences) is stronger under conditions where consumers have to tradeoff two numeric attributes.

### Sample Stimuli



## APPENDIX B3

### *Ratings Skew*

The primary purpose of this study was to test the role of ratings' skew in moderating the joint effect of review volumes and average product ratings. The skew of ratings provides additional information about the expected quality of a product by informing consumers how different proportions of the consumers felt about a product. Khare et al. (2011) demonstrated that ratings' skew affected consumers' judgments only when review volumes were high. However, since Khare et al. investigated single option choices, effect of skew of product ratings was not tested in a setting where consumers have to tradeoff between review volumes and product ratings, the context of this paper. Furthermore, Fisher, Newman, and Dhar (2018) demonstrate a binary bias where consumers neglect differences between more extreme (1- and 5-star) and less extreme (2- and 4-star) valences, so while we only examine three distributions, the exhibited effects would appear to be representative of a vast majority of distributions.

### *Method*

*Participants and design.* We recruited 167 undergraduate students in exchange for course credit. Participants were randomly assigned to a condition in a 2 (review volume levels: low, high) x 3 (option A ratings' skew: absent, negative, positive) between-subjects factorial. The sample size was a convenience sample based on the undergraduate participants for a one-week time period.

*Procedure.* We manipulated review volumes across two product options in the same manner as in prior studies. In addition, the skew of average product ratings was manipulated by reporting the percentage of reviews for each possible rating for Option A

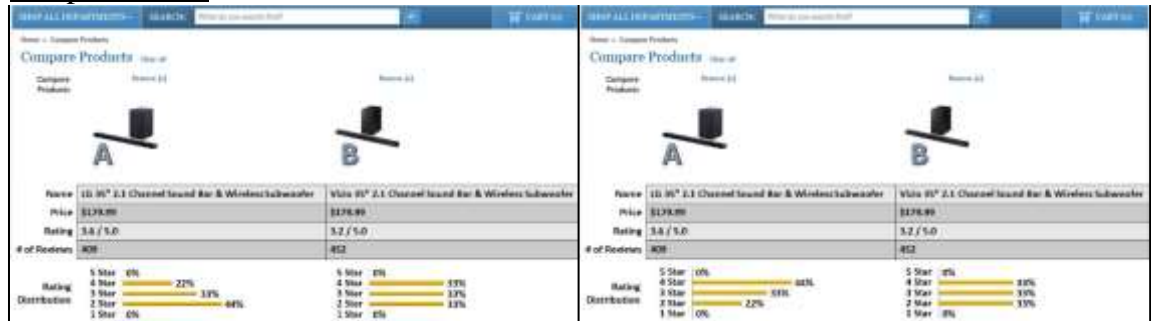
(higher-rated, fewer reviews option). In the positive condition, the percentage of reviews was 0%, 22%, 33%, 44%, 0% for 1-5 stars respectively, demonstrating that they had twice as many 4-star reviews as 2-star reviews. In the negative condition, the pattern was reversed (0%, 44%, 33%, 22%, 0%). For all conditions where skew was present, Option B always had an even dispersion of 0%, 33%, 33%, 33%, 0% for 1-5 stars respectively. Next, participants indicated relative preference between two choice options.

*Results.* A 2 (review volumes) by 3 (skew) ANOVA on relative preference yielded significant main effects of volume ( $F(1, 161) = 16.64; p < .001$ ) and skew ( $F(2, 161) = 26.82; p < .001$ ). The interaction was not significant ( $p > .30$ ). Consistent with prior studies, preference for the higher-rated, fewer reviews option was greater in the presence of high review volumes ( $M_{\text{high}} = 3.30$ ) relative to when the consideration set featured low review volumes ( $M_{\text{low}} = 4.25$ ). Planned contrasts demonstrated that preference for the higher-rated, fewer reviews option was weaker in the presence of a negative skew ( $M_{\text{negative}} = 4.91$ ) relative to when a skew was absent ( $M_{\text{absent}} = 3.62; t(164) = -4.30; p < .001$ ). Furthermore, a positive skew increased preference for the higher-rated, fewer reviews option relative to when a skew was absent ( $M_{\text{positive}} = 2.86; t(164) = -2.53; p = .012$ ).

*Discussion.* This study demonstrated that ratings skew plays a significant role in consumer preferences, but does not interact with the other review attributes. We argue that this occurs because skew is less diagnostic (i.e., more difficult to interpret) relative to average product ratings and review volumes, and thus, consumers are less likely to use it in their judgments relative to the other attributes. Thus, in the presence of other

diagnostic attributes, ratings skew influences consumers but not as a function of review volume levels.

### Sample Stimuli





## APPENDIX B4

### *Aesthetic Products*

This study was designed to demonstrate that review volume effects are attenuated for products which are high in aesthetic value and low in functional value. For such products (e.g., artwork), preference is decided largely on personal tastes rather than the opinions of anonymous others. As such, we expected the effect of review volumes to be null when participants considered artwork.

*Participants and design.* One-hundred and forty-seven participants ( $M_{\text{age}} = 20.91$ ; 50% female; undergraduate sample; course credit) were randomly assigned to one of three review volume levels conditions (low, high, control) in a between-subjects design. The sample size was a convenience sample based on the undergraduate participants for a one-week time period.

*Choice set.* Participants viewed a choice set of four pieces of art, where options were nearly identical with the exception of their average product ratings and review volumes. While one choice option had the highest rating with the fewest reviews (e.g., 3.9, 3 reviews), another choice option had the lowest rating with the most reviews (e.g., 3.1, 55 reviews), and two other choice options in the middle were compromise choice options which were neither the highest, nor lowest on either attribute but were superior on one relative to the other compromise choice option (e.g., 3.7, 17 reviews and 3.3, 41 reviews). Review volume levels were manipulated by either withholding the review volumes in the control condition or adding 200 reviews to the volumes reported above in the high review volume level condition.

*Measures.* To capture preference among the four choice options, we used a discrete choice measure of the highest-rated, fewest reviews choice option, rather than the relative preference measure used in the previous study. To assess the likelihood of choice deferral, we asked participants “Are you more likely to purchase one of the available options or defer purchase, and look elsewhere?” and analyzed this as a binary measure. Lastly, to assess the need for more information, participants were asked “How would you classify the amount of information provided?” on a 7-point scale (1 = not enough information, 7 = too much information). A more difficult tradeoff would require more information to help participants make a decision, thus participants in the low review volume levels condition would be expected to require more information relative to those in the other conditions.

### *Results*

*Choice of the highest-rated, fewest reviews choice option.* A binary logistic regression in which we dummy coded our review volume levels yielded no omnibus effect of review volume levels ( $\chi^2(2) = 1.67; p > .40$ ). As expected, review volume levels did not influence participants’ choice in artwork.

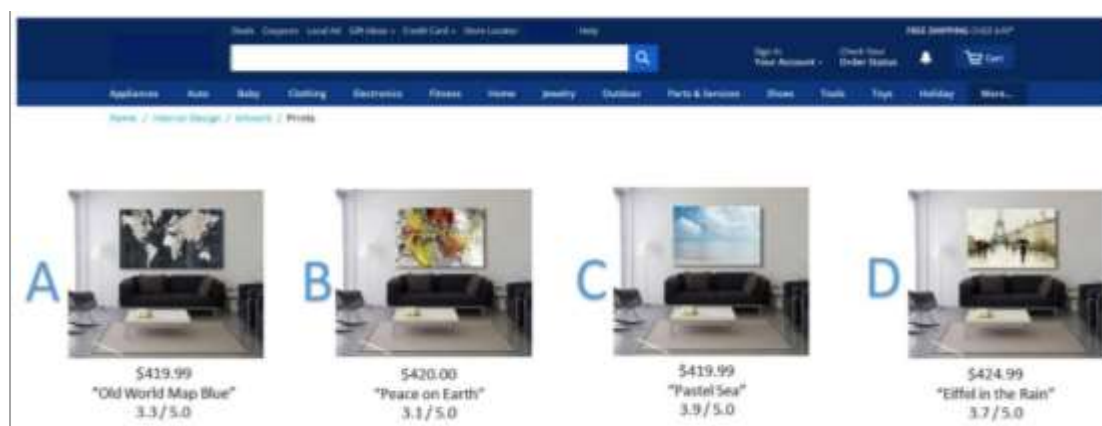
*Rate of choice deferral.* A binary logistic regression in which we dummy coded review volume levels yielded no omnibus effect of review volume levels ( $\chi^2(2) = 1.36; p > .50$ ). As expected, review volume levels did not influence participants’ choice in artwork.

*Need for additional information.* A one-way ANOVA of review volume levels on need for additional information yielded a main effect of review volumes ( $F(2, 144) = 4.49; p = .013$ ). Participants felt they had more information when review volumes were

high ( $M_{\text{high}} = 4.20$ ) relative to absent ( $M_{\text{absent}} = 3.57$ ;  $t(144) = 3.00$ ;  $p = .003$ ). However, the low review volumes led to no significant difference from the other conditions ( $M_{\text{low}} = 3.88$ ; both  $p$ 's  $> .10$ ). Although this finding was unexpected, the need for information appeared to have no effect on choice.

### Discussion

This study provided evidence for the role of review volumes when products were largely a matter of personal taste. When consumers are choosing products whose value is largely aesthetic, they rely on their personal preference to inform their decisions rather than the average product ratings and review volumes. Thus, review volume effects are attenuated when consumers can rely less on the opinion of others and more on their own opinion to shape their decisions. This finding is consistent with a large body of work which has suggested that word-of-mouth is valuable because it helps to reduce uncertainty in consumers' decisions (Brown and Reingen 1987), so when consumers are able to form certain evaluations, they do not need reviews to help aid their decision process.



## APPENDIX B5

### *Single versus Joint Evaluation of Choice Options*

The purpose of this study was to demonstrate the robustness of the interactive effect of review volumes and product ratings on consumer preference under different presentation modes: sequential (i.e., single) versus simultaneous (i.e., joint). Prior research has demonstrated that presentation mode can influence how consumers process attributes, specifically, attenuating numerosity effect (Schley, Lembregts, and Peters 2017).

#### *Method*

*Participants and design.* We recruited 165 participants ( $M_{\text{age}} = 41.09$ ; 55% female) from Amazon mTurk in exchange for a \$0.50 payment. Participants were randomly assigned to a condition in a 2 (review volume levels: low, high) by 2 (presentation mode: sequential, simultaneous) between-subjects by 2 (product replicates: digital camera, tower fan) within-subject mixed design. The sample size was determined based on balancing a 50-subject rule-of-thumb for online samples at the time the study was conducted with cost considerations. The goal was 40 subjects per cell a priori.

*Procedure.* We manipulated review volumes across two product options in the same manner as prior studies. In addition, presentation mode was manipulated by presenting the options either simultaneously (as in prior studies) or one at a time. After viewing both options, participants indicated preference between options and relative importance of the attributes, measured as in Study 6.

#### *Results*

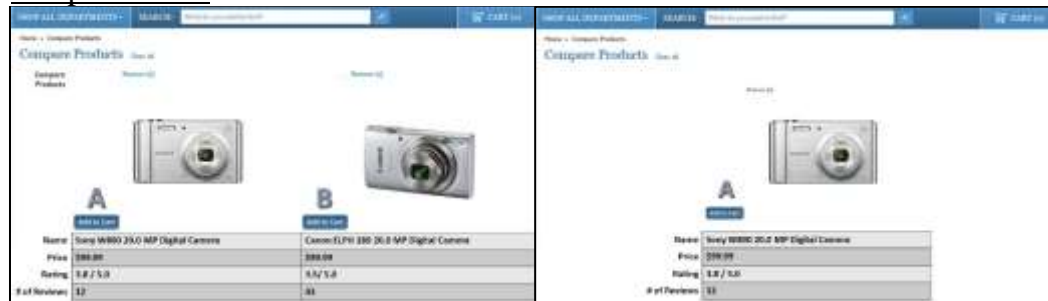
*Option preference.* A repeated measures 2 (review volumes) by 2 (presentation mode) by 2 (product replicates) ANOVA on relative preference yielded significant main effects of volume ( $F(1, 161) = 30.78; p < .001$ ) and product replicates ( $F(1, 161) = 14.93; p < .001$ ). No other effects were significant ( $p > .10$ ). Preference for the higher-rated, fewer reviews option was greater for the tower fan ( $M_{\text{fan}} = 3.46$ ) relative to the digital camera ( $M_{\text{camera}} = 4.16$ ). Consistent with prior studies, preference for the higher-rated, fewer reviews option was greater in the presence of high review volumes ( $M_{\text{high}} = 3.24$ ) relative to when the choice set featured low review volumes ( $M_{\text{low}} = 4.40$ ).

*Mediation via attribute diagnosticity.* As expected, no significant differences existed across image, brand, and price as a function of presentation mode or review volumes ( $p > .25$ ). As before, we calculated the difference in the perceived diagnosticity of average product ratings and review volumes to create a difference measure, averaging across product replicates. A one-way ANOVA on the difference measure yielded a significant effect of review volumes ( $F(1, 161) = 8.29; p = .005$ ). Neither the effects of presentation mode nor their interaction were significant ( $p > .50$ ). Consistent with prior studies, the difference in perceived diagnosticity between the attributes was smaller when review volumes were low relative to high ( $M_{\text{low}} = .30, M_{\text{high}} = .92$ ). Once again, we find support for mediation (PROCESS Model 4; Preacher, Rucker and Hayes 2007) of the effect of review volumes on preference via the difference in perceived diagnosticity of average product ratings and review volumes ( $B = -1.46$ ; 95% confidence interval  $[-.37, -.03]$ ).

*Discussion.* This study provided further support for the robustness of our effect.

Whether consumers evaluate choice options simultaneously or sequentially, the review volumes of considered options play a significant role in the preference between options.

### Sample Stimuli



## APPENDIX B6

### *Popularity Cue*

The purpose of this study was to rule out product popularity as an explanation for the influence of review volumes on consumer preference. If review volumes were solely cues of popularity, we would expect that another popularity cue (e.g., “Best seller”) would attenuate the influence of review volumes on preference.

### *Method*

*Participants and design.* We recruited 402 undergraduate students in exchange for course credit. Participants were randomly assigned to a condition in a 2 (review volume levels: low, high) by 2 (popularity cue: absent, present) between-subjects design. The sample size was a convenience sample based on the undergraduate participants for a two-week time period.

*Procedure.* We manipulated review volumes across two product options in the same manner as prior studies. In addition, in the popularity cue condition, the higher-rated, fewer reviews option had a “Best Seller” badge. Thus, consumers could choose between an option with fewer reviews but a “Best Seller” badge or an option with more reviews without a badge. Next, participants indicated relative preference between two choice options.

*Results.* A 2 (review volumes) by 2 (popularity cue) ANOVA on relative preference yielded significant main effects of volume ( $F(1,398) = 124.43; p < .001$ ) and popularity cue ( $F(1,398) = 14.90; p < .001$ ). The interaction was not significant ( $p > .20$ ). Consistent with prior studies, preference for the higher-rated, fewer reviews option was lower in the presence of low versus high review volumes ( $M_{\text{low}} = 4.63$  vs.  $M_{\text{high}} = 2.74$ ).

Furthermore, preference for the higher-rated, fewer reviews option was also greater when that option was labeled as a “Best Seller” ( $M_{\text{present}} = 3.34$  vs.  $M_{\text{absent}} = 4.01$ ).

*Discussion.* This study ruled out popularity as an alternative explanation. It demonstrated that when consumers are choosing from “Best Sellers”, review volumes still play a critical role in the decision process when those best sellers also have low review volumes. It also did demonstrate a main effect of the presence of the badge such that the mere presence of best-selling options increased preference for the higher-rated one, suggesting that the badge acts as an additional discriminating cue in judgment processes.

### Sample Stimuli





## APPENDIX B7

### *Product Versions*

The purpose of this study was to demonstrate the persistent effect of review volumes in light of possible justification for low review volumes. In this case, we used product versions (e.g., 2015 vs. 2013 models). Because newer versions have been on the market for less time, their lower review volumes should be justified. Yet, even in this case, we argue that the product version will not attenuate the effect of review volumes, demonstrating that even with proper justification, low review volumes still significantly influence consumer preference.

### *Method*

*Participants and design.* We recruited 152 participants ( $M_{\text{age}} = 36.43$ ; 52% female; Amazon mTurk; \$0.50 payment). Participants were randomly assigned to a condition in a 2 (review volumes: low, high) by 2 (product age information: absent, present) between-subjects by 2 (product replicates: DVD player, phone) within-subject mixed design. The sample size was determined based on a 40-subject rule-of-thumb for online samples at the time the study was conducted, however not all participants completed the study.

*Procedure.* We manipulated review volumes across two product options in the same manner as prior studies. In addition, product age information was manipulated slightly differently for each of the product replicated. For the DVD player, we labeled each option with a release year: Option A was a 2015 version while Option B was a 2013 version. For the cell phone, we labeled each option with a version number: Option A was a Galaxy S7 while Option B was a Galaxy S6. Thus, each choice set included a higher-

rated, fewer reviews, newer version option and its inverse. For each product replicate, participants indicated relative preference between choice options.

*Results.* A repeated measures 2 (review volumes) by 2 (product age information) by 2 (product replicate) ANOVA on relative preference yielded significant main effects of review volume ( $F(1, 148) = 22.59; p < .001$ ) and product age ( $F(1, 148) = 8.91; p = .003$ ). Neither the effect of product replicate nor the interactions were significant ( $p > .15$ ). Consistent with prior studies, preference for the higher-rated, fewer reviews option was lower in the presence of low vs. high review volumes ( $M_{\text{low}} = 3.28$  vs.  $M_{\text{high}} = 2.29$ ). Furthermore, preference for the higher-rated, fewer reviews option was also greater when it was a newer product ( $M_{\text{new product}} = 2.45$  vs.  $M_{\text{no product age information}} = 3.09$ ).

*Discussion.* This study demonstrated the persistent influence of low review volumes even when a low number of reviews is justified by being a newer product. While a main effect existed such that preference for the higher-rated, fewer reviews option was greater when the product's newer version was disclosed versus when it was not, this disclosure of version did not attenuate the influence of review volumes. As such, this study demonstrated that consumers can be led to prefer older products simply because they also have more reviews. Given the rapid advances in technology, this study provides some evidence that consumers may make suboptimal decisions as a function of review volumes.

### Sample Stimuli

WAP ALL DEPARTMENTS SEARCH What do you want to buy? CART 0/1

Home > Compare Products

### Compare Products

View all Compare Products Reviews (4)



**A**  
Add to Cart



**B**  
Add to Cart

Name:	DBPower 9.5" Portable DVD Player 2015	DBPower 9.5" Portable DVD Player 2013
Price:	\$65.99	\$69.99
Rating:	3.8 / 5.0	3.5 / 5.0
# of Reviews:	14	9

WAP ALL DEPARTMENTS SEARCH What do you want to buy? CART 0/1

Home > Compare Products

### Compare Products

View all Compare Products Reviews (2)



**A**  
Add to Cart



**B**  
Add to Cart

Name:	Samsung Galaxy S7 (certified refurbished)	Samsung Galaxy S6 (certified refurbished)
Price:	\$129.99	\$139.99
Rating:	5.0 / 5.0	5.3 / 5.0
# of Reviews:	15	41

## APPENDIX B8

### *New Product Arrival*

One reason that a choice may have a higher average product rating and fewer reviews is because it is a newer choice option relative to the competition. As such, it has fewer reviews but better quality (for example, newer technology) leading to a higher average product rating. Next, we test whether introducing new arrival cue will moderate the influence of review volumes on preferences.

### *Method*

*Participants and design.* We recruited 202 participants from mTurk ( $M_{\text{age}} = 36.37$ ; 43% female) in exchange for a \$0.50 payment. Participants were randomly assigned to a condition in a 2 (review volume levels: low, high) x 2 (new arrival cue: absent, present) between-subjects design by 3 (product replicates: tower fan, cookie sheets, knife sets) within-subject, mixed design. The sample size was determined based on a 50-subject rule-of-thumb for online samples at the time the study was conducted.

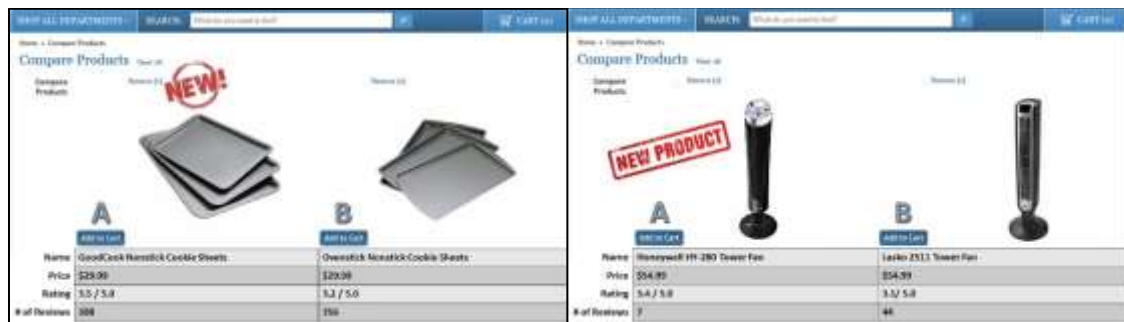
*Procedure.* We manipulated review volumes across two product options in the same manner as prior studies. In the new arrival cue present condition, Option A also had a badge which denoted that the product was “new”. We used slightly different badges (e.g., “new”, “new arrival”, and “new product”) across the three product categories to improve generalizability. For each product replicate, participants indicated relative preference between the two choice options.

*Results.* A 2 (review volumes) by 2 (new arrival cue) by 3 (product replicates) repeated-measures ANOVA on relative preference yielded only significant main effect of volume ( $F(1, 198) = 20.83$ ;  $p < .001$ ). No other effects were significant ( $p > .15$ ).

Consistent with prior studies, preference for the higher-rated, fewer reviews option was greater in the presence of high review volumes ( $M_{\text{high}} = 2.87$ ) relative to when the consideration set featured low review volumes ( $M_{\text{low}} = 3.72$ ).

*Discussion.* This study demonstrated the robustness of our effect in light of a clear explanation for the low review volume: products being labeled as “new”. Thus, consumers are still likely to demonstrate weakened preference for higher-rated options merely because they have fewer reviews, even when they are newer, a justification for why they would have fewer reviews.

### Sample Stimuli



## APPENDIX B9

### *Reviewer Credibility*

The purpose of this study was to demonstrate the effect of review credibility on how consumers incorporate average product ratings and review volumes into their decisions.

### *Method*

*Participants and design.* We recruited 226 undergraduate participants in exchange for course credit. Participants were randomly assigned to one of two review volume levels condition (low, high) crossed with one of two credibility conditions (absent, present) between-subjects. The sample size was a convenience sample based on the undergraduate participants for a one-week time period.

*Procedure.* We manipulated review volumes in the same manner as prior studies. Credibility was manipulated by the presence of a “Consumer Reports Verified” badge for option A, (higher rated, fewer reviews option). After viewing both options, participants were given the choice between Option A, Option B, or to Defer Purchase. Then participants were asked the importance of each attribute.

### *Results*

*Option preference.* A 2 (review volume levels: low, high) x 2 (credibility cue: present, absent) binary logit yielded a main effect of review volume levels ( $B = -2.90$ ; Wald  $\chi^2 = 7.88$ ;  $p = .005$ ) on option choice. Neither the effect of credibility nor the interaction were significant ( $p > .70$ ). Consistent with prior studies, preference for the higher-rated, fewer reviews option was greater in the presence of high review volumes ( $P_{\text{high}} = 93\%$ ) relative to when the choice set featured low review volumes ( $P_{\text{low}} = 52\%$ ).

*Choice Deferral.* A 2 (review volume levels: low, high) x 2 (credibility: present, absent) binary logit yielded no significant effects on deferral rates ( $p > .20$ ), though it was directionally consistent with prior studies where the rate of deferral was greater when review volumes were low ( $P_{\text{low}} = .13$ ) relative to high ( $P_{\text{high}} = .05$ ).

*Mediation via attribute diagnosticity.* As expected, no significant differences existed across image, brand, and price as a function of presentation mode or review volumes ( $p > .15$ ). As before, we calculated the difference in the perceived diagnosticity of average product ratings and review volumes to create a difference measure. A one-way ANOVA on the difference measure yielded a significant effect of review volumes ( $F(1, 213) = 11.26; p = .001$ ). Consistent with prior studies, the difference in perceived diagnosticity between the attributes was smaller when review volumes were low relative to high ( $M_{\text{low}} = .50, M_{\text{high}} = 1.01$ ). Once again, we find support for mediation (PROCESS Model 4; Preacher, Rucker and Hayes 2007) of the effect of review volumes on preference via the difference in perceived diagnosticity of average product ratings and review volumes ( $B = -.41$ ; 95% confidence interval  $[-.86, -.14]$ ).

*Discussion.* This study provided further evidence of previous findings while also ruling out review credibility as a potential moderator. Whether the reviews are aggregated from the general population or verified by Consumer Reports, the influence of low review volumes persists. This study did not demonstrate an effect of review volume levels on the rates of deferral. We posit that this could be an effect of the badge as it mitigates the risk from fewer reviews, but not enough to completely shift preference to that option.

#### Sample Stimuli

VIEW ALL PRODUCTS		10/20/16	10/20/16	10/20/16
Compare Products		Compare Products		
				
<b>A</b> 		<b>B</b> 		
<b>Customer Reviews</b> <b>Verified</b>		<b>Customer Reviews</b> <b>Verified</b>		
<b>Name</b>		<b>Name</b>		
Cuisinart® 1.5 Liter Convection Toaster Oven		Hamilton Beach® Convection Toaster Oven		
<b>Price</b>		<b>Price</b>		
\$109.99		\$109.99		
<b>Rating</b>		<b>Rating</b>		
3.7 / 5.0		3.4 / 5.0		
<b># of Reviews</b>		<b># of Reviews</b>		
100		100		



## APPENDIX C – ESSAY II STIMULI

### *Appendix C1 - Study 2 Stimuli*

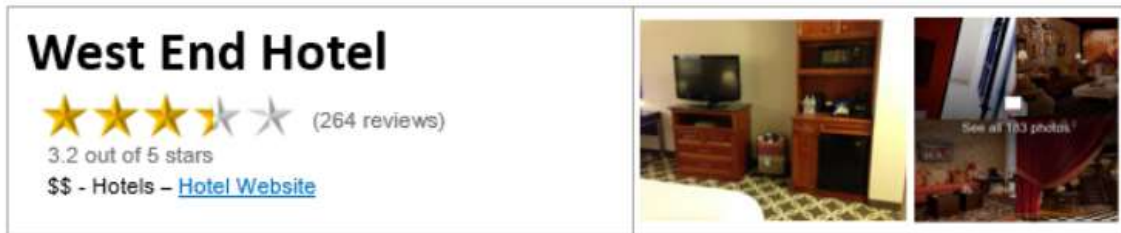
#### *[Round 1]*

Please imagine the following situation:

You have an upcoming weekend free and decide to take a mini-vacation to San Diego, California. Since you aren't familiar with the area, you don't have any information on where to stay. So you decide to check out some online reviews to find out where to stay.

While searching for hotels on Yelp, you find a hotel that seems to have a great location and fits your budget perfectly, so you decide to click on the "West End Hotel" to learn more about it.

Before reaching the reviews for "West End Hotel", you see the following information.



After viewing this, you decide to inquire further and check out some reviews.

Recent reviews for the hotel on the review website are shown below. Please take some time to look over them now.

---

## RECOMMENDED REVIEWS

**Linda P.**

**Rating:** ★★☆☆☆

SHARE THIS REVIEW

- Share on Google+
- Share on Facebook
- Share on Twitter
- Email to a friend

The hotel is situated on a bustling/lively street. It has seen better days and somewhat run down. Although the room ...

[Read more ▼](#)

**Elizabeth J.**

**Rating:** ★☆☆☆☆

SHARE THIS REVIEW

- Share on Google+
- Share on Facebook
- Share on Twitter
- Email to a friend

Travel for business often and this place is horrible! Room was musty, balcony in the lobby had a pane of glass ...

[Read more ▼](#)

**James C.**

**Rating:** ★★★★★

SHARE THIS REVIEW

- Share on Google+
- Share on Facebook
- Share on Twitter
- Email to a friend

We just spent six nights at this lovely, clean, homey and reasonable place. Our room was large, comfortable and ...

[Read more ▼](#)

**Barbara L.**

**Rating:** ★★☆☆☆

SHARE THIS REVIEW

- Share on Google +
- Share on Facebook
- Share on Twitter
- Email to a friend

This hotel sounds lovely online...but beware! The room was dirty, the light did not work (and it was never fixed!), ...

[Read more ▾](#)

**Dan H.**

**Rating:** ★★★★★

SHARE THIS REVIEW

- Share on Google +
- Share on Facebook
- Share on Twitter
- Email to a friend

I have stayed at this property on numerous occasions and have never had a problem. My room is always nice and ...

[Read more ▾](#)

**Michael F.**

**Rating:** ★★☆☆☆

SHARE THIS REVIEW

- Share on Google +
- Share on Facebook
- Share on Twitter
- Email to a friend

Stayed two nights—one day before my two week cruise and one day after. I arrived early both times but could not ...

[Read more ▾](#)

**Patricia K.**

**Rating:** ★★★★★

SHARE THIS REVIEW

- Share on Google +
- Share on Facebook
- Share on Twitter
- Email to a friend

We found the rooms spacious and comfortable. The breakfast was better than expected and sitting on the ...

[Read more ▾](#)

**William M.**

**Rating:** ★☆☆☆☆

SHARE THIS REVIEW  
[Share on Google +](#)  
[Share on Facebook](#)  
[Share on Twitter](#)  
[Email to a friend](#)

This hotel sucks. I stayed here with my family and from when we arrived the staff was not friendly, the room I ...

[Read more ▼](#)

**Mary B.**

**Rating:** ★★★★★

SHARE THIS REVIEW  
[Share on Google +](#)  
[Share on Facebook](#)  
[Share on Twitter](#)  
[Email to a friend](#)

Went on a last minute whim- GREAT experience- Rooms have just been re-decorated - staff was VERY friendly ...

[Read more ▼](#)

**Robert S.**

**Rating:** ★★★★☆

SHARE THIS REVIEW  
[Share on Google +](#)  
[Share on Facebook](#)  
[Share on Twitter](#)  
[Email to a friend](#)

Needed a room for one night in San Diego. Used my reward points as the hotel was charging an incredible ...

[Read more ▼](#)

## NOT RECOMMENDED

(We use automated software to recommend the reviews we think will be the most helpful to the Yelp community based primarily on quality, reliability and the reviewer's activity on Yelp. Advertisers get no special treatment. The reviews below didn't make the cut and are therefore not factored into this business's overall star rating. Check out our FAQ for more details. )

**Kyle.**

**Rating:** ★★★★☆

SHARE THIS REVIEW  
[Share on Google +](#)  
[Share on Facebook](#)  
[Share on Twitter](#)  
[Email to a friend](#)

Decided to try this place out since it seemed like a good deal. After checking in, I took a dip in the pool, which was ...

[Read more ▼](#)

Which review would you like to read in full?

Linda P. (★★★★☆)

Elizabeth J. (★☆☆☆☆)

James C. (★★★★★)

Barbara L. (★★☆☆☆)

Dan H. (★★★★☆)

Michael F. (★★☆☆☆)

Patricia K. (★★★★☆)

William M. (★☆☆☆☆)

Mary B. (★★★★★)

Robert S. (★★★★☆)

Kyle K. (★★★★☆)

---

**Elizabeth J.**

**Rating:** ★☆☆☆☆

SHARE THIS REVIEW

-  Share on Google +
-  Share on Facebook
-  Share on Twitter
-  Email to a friend

Travel for business often and this place is horrible! Room was musty, balcony in the lobby had a pane of glass broken and had yellow "caution" tape up to prevent a fall from second story. Very noisy hotel- not for business travelers. Hotel management was the most unpleasant, unkind and disrespectful people I have ever met. I am surprised they are in business with such an attitude about customer service! Desk staff was very nice and helpful.

---

**William M.**

**Rating:** ★☆☆☆☆

SHARE THIS REVIEW

-  Share on Google +
-  Share on Facebook
-  Share on Twitter
-  Email to a friend

This hotel sucks. I stayed here with my family and from when we arrived the staff was not friendly, the room I ordered was non smoking but it was filled with smoke and then when they put me in a new room it was the same. Don't waste your money stay somewhere else.

---

**Barbara L.**

**Rating:** ★★☆☆☆

SHARE THIS REVIEW

-  Share on Google +
-  Share on Facebook
-  Share on Twitter
-  Email to a friend

This hotel sounds lovely online...but beware! The room was dirty, the light did not work (and it was never fixed!), and the food was pretty lame too! It is not a very nice area too... A lot of undesirables on the street... and almost makes it a bit frightening to go out... On the good side...its handy for Sea World... but that is the only good side!

---

**Michael F.**

**Rating:** ★★☆☆☆

SHARE THIS REVIEW

- Share on Google +
- Share on Facebook
- Share on Twitter
- Email to a friend

Stayed two nights--one day before my two week cruise and one day after. I arrived early both times but could not check in until pretty near the normal 3:00pm check in time both days as the rooms were not cleaned yet. On the plus side, there are a couple of excellent restaurants nearby to relax in. Also, on both nights the evening front desk staff had a somewhat rude, non-friendly attitude when asked about shuttle transportation and other general questions.

---

**Linda P.**

**Rating:** ★★★☆☆

SHARE THIS REVIEW

- Share on Google +
- Share on Facebook
- Share on Twitter
- Email to a friend

The hotel is situated on a bustling/lively street. It has seen better days and somewhat run down. Although the room was renovated, the hotel looked tired. When we arrived on Saturday evening, we enjoyed some loud and lively music from a bar in the street which was not ideal for our young children.

---

**Robert S.**

**Rating:** ★★★☆☆

SHARE THIS REVIEW

- Share on Google +
- Share on Facebook
- Share on Twitter
- Email to a friend

Needed a room for one night in San Diego. Used my reward points as the hotel was charging an incredible amount for a regular room. We have stayed here before. The beds are decent but the rooms are very small. Breakfast is decent. There is a tiki bar in the center (outside) and it was incredibly noisy all night. People standing outside our room talking until 5am.



---

**Dan H.**

**Rating:** ★★★★★

SHARE THIS REVIEW

- Share on Google +
- Share on Facebook
- Share on Twitter
- Email to a friend

I have stayed at this property on numerous occasions and have never had a problem. My room is always nice and clean and the parking fee is not outrageous like some hotels. There are lots of places to dine and hang out within walking distance. The staff as a whole are nice. I would recommend this hotel.

---

**Patricia K.**

**Rating:** ★★★★★

SHARE THIS REVIEW

- Share on Google +
- Share on Facebook
- Share on Twitter
- Email to a friend

We found the rooms spacious and comfortable. The breakfast was better than expected and sitting on the patio enjoying the morning sunshine was great. The best thing for us was the service we received from two of the receptionists Alex and Yohanka. A big thank you to these two marvellous people.

---

**James C.**

**Rating:** ★★★★★

SHARE THIS REVIEW

- Share on Google +
- Share on Facebook
- Share on Twitter
- Email to a friend

We just spent six nights at this lovely, clean, homey and reasonable place. Our room was large, comfortable and impeccably clean, the free breakfast was delicious, the staff was lovely and the pool was absolutely delightful. We felt completely at home and really appreciated the easiness and homelike atmosphere. We enjoyed being downtown and yet felt that we were in an oasis. Great job, everyone, who works so hard to keep this hotel so nicely run!



---

**Mary B.**

Rating: ★★★★★

SHARE THIS REVIEW

[Share on Google +](#)  
[Share on Facebook](#)  
[Share on Twitter](#)  
[Email to a friend](#)

Went on a last minute whim- GREAT experience- Rooms have just been re-decorated - staff was VERY friendly and helpful and the location was PERFECT!!! You can walk to most attractions and if the Sea World is on your list, this is the perfect spot for convenience. Best place in San Diego!!

---

**Kyle K.**

Rating: ★★★★★

SHARE THIS REVIEW

[Share on Google +](#)  
[Share on Facebook](#)  
[Share on Twitter](#)  
[Email to a friend](#)

Decided to try this place out since it seemed like a good deal. After checking in, I took a dip in the pool, which was okay. Then I ordered room service which was decent. Once the night set, I ventured outside to see what to do. There were a few options, it wasn't too hard to choose. You could do worse, and you could do better, if you were coming to the area.

---

Which review would you like to read next?

- ☐ Linda P. (★★★★☆)
- ☐ Elizabeth J. (★☆☆☆☆)
- ☐ James C. (★★★★★)
- ☐ Barbara L. (★★☆☆☆)
- ☐ Dan H. (★★★★☆)
- ☐ Michael F. (★★☆☆☆)
- ☐ Patricia K. (★★★★☆)
- ☐ William M. (★☆☆☆☆)
- ☐ Mary B. (★★★★★)
- ☐ Robert S. (★★★★☆)
- ☐ Kyle K. (★★★★☆)
- ☐ I'm finished reading reviews

---

Now that you have finished reading the reviews, we have some questions about the hotel.

---

Based on the information you have seen, what star rating would you give this hotel? That is, how many stars do you think truly reflects the quality of West End Hotel?

My rating: ★★☆☆☆

---

To what degree would you say that your rating of West End Hotel is accurate?

Not at all                      Extremely

☐      ☐      ☐      ☐      ☐      ☐      ☐

---

To what degree would you say that your rating of West End Hotel was made to reward the business?

Not at all                      Extremely

☐      ☐      ☐      ☐      ☐      ☐      ☐

---

To what degree would you say that your rating of West End Hotel was made to punish the business?

Not at all ☐ ☐ ☐ ☐ ☐ ☐ Extremely ☐

---

Given the available information, how accurate is the 3.2 out of 5.0 average star rating from other consumers for West End Hotel?

Not at all ☐ ☐ ☐ ☐ ☐ ☐ Extremely ☐

---

Now we have a few additional questions about West End Hotel.

How strongly would you consider staying at this hotel?

I would definitely not consider staying here ☐ ☐ ☐ ☐ ☐ I would strongly consider staying here ☐

---

How likely would you be to suggest this hotel to a friend that's also visiting the town?

Not at all likely      Extremely likely

☐      ☐      ☐      ☐      ☐      ☐

---

How much would you be willing to pay for a room at West End Hotel relative to a traditional chain (e.g., Best Western, Holiday Inn, etc.)?

Much Less      About the Same      Much More

☐      ☐      ☐      ☐      ☐      ☐

☐

---

How much risk would be associated with staying at West End Hotel?

Extremely Low Risk      Extremely High Risk

☐      ☐      ☐      ☐      ☐      ☐

☐

---

What is your perception of West End Hotel's quality?

Extremely Low Quality      Extremely High Quality

☐      ☐      ☐      ☐      ☐      ☐

☐

---

What is your value perception of West End Hotel?

Extremely Bad Value      Extremely Good Value

☐      ☐      ☐      ☐      ☐      ☐

☐

---

Now we'd like you to think about the reviews you read.

How likely is it that the reviews are trustworthy?

Not at all ☐ ☐ ☐ ☐ ☐ ☐ Extremely ☐

---

How likely is it that the reviews are honest?

Not at all ☐ ☐ ☐ ☐ ☐ ☐ Extremely ☐

---

How likely is it that the reviews are accurate?

Not at all ☐ ☐ ☐ ☐ ☐ ☐ Extremely ☐

---

Have you ever encountered a message disclosing that fraudulent reviews have been identified before?

- ☐ No  
☐ Yes
- 

How familiar are you with each of the following:

	Not at all				Extremely
West End Hotel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Yelp.com	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

---

How often do you:

	Never				Often
stay at West End Hotel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
visit Yelp.com	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

---

Please indicate your gender.

- ☐ Male  
☐ Female
- 

Please indicate what your age **will be on December 31, 2017**.

---

Please enter the **LAST 5** digits of your student ID (you may pull out your ID if necessary)

[Round 2]


Please imagine the following situation:

You once again have an upcoming weekend free and decide to take a mini-vacation to San Diego, California. You ended up not going last time, but have decided to once again check out some online reviews for where to stay.

While searching for hotels on Yelp, you revisit the reviews for "West End Hotel" which last time seemed to have a great location and fit your budget perfectly.


Before reaching the reviews for "West End Hotel", you see the following information.


## West End Hotel

(264 reviews)

3.2 out of 5 stars

\$\$ - Hotels – [Hotel Website](#)

[See all 183 photos](#)

**Alert**

Yelp caught a competitor paying others to write negative reviews for this business. This means that someone was trying to artificially decrease the ratings of this business. We weren't fooled, but wanted you to know because these actions not only hurt consumers, but also honest businesses who play by the rules.

After viewing this, you decide to inquire further and check out some reviews.

Recent reviews for the hotel on the review website are shown below. Please take some time to look over them now.

## West End Hotel



3.2 out of 5 stars

\$\$ - Hotels - [Hotel Website](#)



### Alert



Yelp caught this business paying others to write positive reviews for this business. This means that someone was trying to artificially increase the ratings of this business. We weren't fooled, but wanted you to know because these actions not only hurt consumers, but also honest businesses who play by the rules.

### Ryan S.

Rating: ★★★★★

SHARE THIS REVIEW

- Share on Google +
- Share on Facebook
- Share on Twitter
- Email to a friend

I found this hotel on a discount website, and I got what I paid for. A decent place that served its purpose as a

...

[Read more ▼](#)

### Victoria H.

Rating: ★☆☆☆☆

SHARE THIS REVIEW

- Share on Google +
- Share on Facebook
- Share on Twitter
- Email to a friend

Don't bother staying here if you want a decent hotel. Its noisy, dirty, and the pool was filled with some leaves that

...

[Read more ▼](#)

### Timothy B.

Rating: ★★★★★

SHARE THIS REVIEW

- Share on Google +
- Share on Facebook
- Share on Twitter
- Email to a friend

Glad I stayed here. I left some of my toiletries at home, and when I let the front desk know, they delivered some

...

[Read more ▼](#)

### Monica R.

Rating: ★★☆☆☆

SHARE THIS REVIEW

- Share on Google +
- Share on Facebook
- Share on Twitter
- Email to a friend

I was kinda disappointed in this place. I thought it would be a little bit nicer, but it wasn't terrible. At least the staff

...

[Read more ▼](#)



**Peter P.**

Rating: ★★★★★

SHARE THIS REVIEW

 Share on Google +

 Share on Facebook

 Share on Twitter

 Email to a friend

If you're looking for a decent hotel that costs a bit extra for a great location, this is it. Its super conveniently

...


[Read more ▾](#)

**Brian T.**

Rating: ★★★★★

SHARE THIS REVIEW

 Share on Google +

 Share on Facebook

 Share on Twitter

 Email to a friend

This place needs to do better. My room was missing towels when I showed up, and it took them nearly an

...

[Read more ▾](#)

**Diane T.**

Rating: ★★★★★

SHARE THIS REVIEW

📧 Share on Google +

📘 Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

This is a nice hotel that makes for a very relaxing trip. They have good food when you don't want to go out, and

...

[Read more ▾](#)

**Elliott M.**

Rating: ★☆☆☆☆

SHARE THIS REVIEW

📧 Share on Google +

📘 Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

I usually visit this city every year and try different hotels. This place was definitely my least favorite. Its

...

[Read more ▾](#)

**Brittany C.**

Rating: ★★★★★

SHARE THIS REVIEW

📧 Share on Google +

📘 Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

We just spent a full week at this lovely hotel. The staff was great, and super attentive. When we came back every

...

[Read more ▾](#)

**Lyle A.**

Rating: ★★★★★

SHARE THIS REVIEW

📧 Share on Google +

📘 Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

Overall, I'd describe this place as "average". There isn't anything that's going to excite you, but also nothing that

...

[Read more ▾](#)

## NOT RECOMMENDED

*(We use automated software to recommend the reviews we think will be the most helpful to the Yelp community based primarily on quality, reliability and the reviewer's activity on Yelp. Advertisers get no special treatment. The reviews below didn't make the cut and are therefore not factored into this business's overall star rating. Check out our FAQ for more details. )*

**Arnold W.** Rating: ★☆☆☆☆

SHARE THIS REVIEW

 Share on Google +  
 Share on Facebook  
 Share on Twitter  
 Email to a friend

Man, I couldn't be more sad about my choice to stay here. It was a really bad experience from checking in  
...  
[Read more ▼](#)

**Arnold W.** Rating: ★★★★★

SHARE THIS REVIEW

 Share on Google +  
 Share on Facebook  
 Share on Twitter  
 Email to a friend

Man, I couldn't be more happy about my choice to stay here. It was a really good experience from  
...  
[Read more ▼](#)

Which review would you like to read in full?

- ☐ Ryan S. (★★★★☆)
- ☐ Victoria H. (★☆☆☆☆)
- ☐ Timothy B. (★★★★★)
- ☐ Monica R. (★★☆☆☆)
- ☐ Peter P. (★★★★☆)
- ☐ Brian T. (★★☆☆☆)
- ☐ Diane T. (★★★★☆)
- ☐ Elliott M. (★☆☆☆☆)
- ☐ Brittany C. (★★★★★)
- ☐ Lyle A. (★★★★☆)
- ☐ Arnold W. (★☆☆☆☆)
- ☐ Arnold W. (★★★★★)

**Victoria H.**

Rating: ★☆☆☆☆

SHARE THIS REVIEW

📧 Share on Google +

📘 Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

Don't bother staying here if you want a decent hotel. Its noisy, dirty, and the pool was filled with some leaves that had blown in from trees. I ended up having to Uber to everything I wanted to do in the city too.

**Elliott M.**

Rating: ★☆☆☆☆

SHARE THIS REVIEW

📧 Share on Google +

📘 Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

I usually visit this city every year and try different hotels. This place was definitely my least favorite. Its ratty and old, and the people are slow and annoying. I definitely won't be coming back, and don't recommend this spot to anyone.

**Monica R.**

Rating: ★★☆☆☆

SHARE THIS REVIEW

📧 Share on Google +

📘 Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

I was kinda disappointed in this place. I thought it would be a little bit nicer, but it wasn't terrible. At least the staff was nice, even if they were a bit slow. If you're price-conscious, this hotel may be okay, but if you can spend a bit more, I think another place might be better.

**Brian T.**

Rating: ★★☆☆☆

SHARE THIS REVIEW

📧 Share on Google +

📘 Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

This place needs to do better. My room was missing towels when I showed up, and it took them nearly an hour to deliver some towels. At least, they were clean when I finally got them. The front desk folks were nice but a little too slow for my liking.

**Ryan S.**

Rating: ★★★☆☆

SHARE THIS REVIEW

📧 Share on Google +

📘 Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

I found this hotel on a discount website, and I got what I paid for. A decent place that served its purpose as a functional place to sleep and shower. Don't really have anything to complain about, so if you just need a simple, standard hotel, this is an acceptable spot.

**Lyle A.**

Rating: ★★★☆☆

SHARE THIS REVIEW

📧 Share on Google +

📘 Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

Overall, I'd describe this place as "average". There isn't anything that's going to excite you, but also nothing that is going to upset you. If you just need a place to lay your head for the evening, you could do worse, and you could probably do better.

**Peter P.**

Rating: ★★★★★

SHARE THIS REVIEW

📧 Share on Google +

📘 Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

If you're looking for a decent hotel that costs a bit extra for a great location, this is it. Its super conveniently located near some great restaurants and attractions, and the staff is a good bunch. It won't change your life, but it avoids any of the negatives associated with hotels.



**Diane T.**

Rating: ★★★★★

SHARE THIS REVIEW

- 📧 Share on Google +
- 📘 Share on Facebook
- 🐦 Share on Twitter
- ✉ Email to a friend

This is a nice hotel that makes for a very relaxing trip. They have good food when you don't want to go out, and a pool for when you want to relax. Nice staff and clean rooms for a good price makes this a great option.

**Timothy B.**

Rating: ★★★★★

SHARE THIS REVIEW

- 📧 Share on Google +
- 📘 Share on Facebook
- 🐦 Share on Twitter
- ✉ Email to a friend

Glad I stayed here. I left some of my toiletries at home, and when I let the front desk know, they delivered some replacements right to me. They also had great recommendations for local food spots and nightlife. Can't believe I lucked out with this place.

**Brittany C.**

Rating: ★★★★★

SHARE THIS REVIEW

- 📧 Share on Google +
- 📘 Share on Facebook
- 🐦 Share on Twitter
- ✉ Email to a friend

We just spent a full week at this lovely hotel. The staff was great, and super attentive. When we came back every evening, our room looked like it was brand new, and they left a nice chocolate on our pillows! If we ever come back to this area, we'll definitely consider staying here first.

**Arnold W.**

Rating: ★☆☆☆☆

SHARE THIS REVIEW

- 📧 Share on Google +
- 📘 Share on Facebook
- 🐦 Share on Twitter
- ✉ Email to a friend

Man, I couldn't be more sad about my choice to stay here. It was a really bad experience from checking in (waiting forever) to leaving (being overpaid), and all that bad stuff in between: dirty room, noisy neighbor, small windows, etc.

**Arnold W.**

Rating: ★★★★★

SHARE THIS REVIEW

g+ Share on Google +

f Share on Facebook

Share on Twitter

Email to a friend

Man, I couldn't be more happy about my choice to stay here. It was a really good experience from checking in (almost immediately) to leaving (receiving a discount), and all that good stuff in between: clean room, quiet neighbor, large windows, etc.

Which review would you like to read next?

- ☐ Ryan S. (★★★★☆)
- ☐ Victoria H. (★☆☆☆☆)
- ☐ Timothy B. (★★★★★)
- ☐ Monica R. (★★☆☆☆)
- ☐ Peter P. (★★★★☆)
- ☐ Brian T. (★★☆☆☆)
- ☐ Diane T. (★★★★☆)
- ☐ Elliott M. (★☆☆☆☆)
- ☐ Brittany C. (★★★★★)
- ☐ Lyle A. (★★★★☆)
- ☐ Arnold W. (★☆☆☆☆)
- ☐ *I'm finished reading reviews*
- ☐ Arnold W. (★★★★★)

[same measures as Round 1]

*[Round 3]*

Please imagine the following situation:

Your upcoming free weekend is almost here and you really need to book a hotel. So you decide to revisit the hotel on Yelp that you noticed a few weeks ago and wanted to check it out one last time before deciding whether to stay there. After going to Yelp, you search for the "West End Hotel" and read some reviews once more.

[same reviews and measures as Round 1]

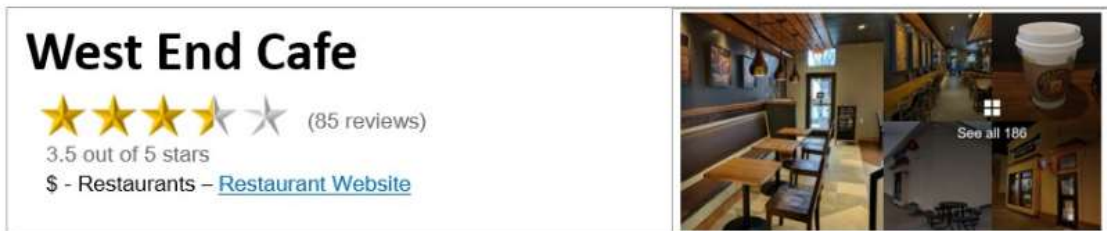


### Appendix C2 – Study 3 Stimuli

Please imagine the following situation:

While on a short trip, you decide to check Yelp for a cafe where you could get something to drink and a bite to eat. While browsing the various cafes, one in particular, "West End Cafe" catches your eye. On the next page, you'll view the aggregate information about West End Cafe.

At the top of the page for "West End Cafe", you see the following information.



After viewing this, you decide to inquire further and check out some reviews.

## West End Cafe



3.5 out of 5 stars

\$ - Restaurants – [Restaurant Website](#)



### Alert



Yelp caught a competitor paying others to write negative reviews for this business. This means that someone was trying to artificially decrease the ratings of this business. We weren't fooled, but wanted you to know because these actions not only hurt consumers, but also honest businesses who play by the rules.

## West End Cafe



3.5 out of 5 stars

\$ - Restaurants – [Restaurant Website](#)



### Alert



Yelp caught this business paying others to write positive reviews for this business. This means that someone was trying to artificially increase the ratings of this business. We weren't fooled, but wanted you to know because these actions not only hurt consumers, but also honest businesses who play by the rules.

## RECOMMENDED REVIEWS

**Rose S.**

Rating: ★★★★★

SHARE THIS REVIEW

📧 Share on Google +

f Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

I usually come here to get a sandwich when I'm in the area and hungry. I've tried a bunch of stuff on the menu

...

[Read more ▾](#)

**Dan T.**

Rating: ★★★★★

SHARE THIS REVIEW

📧 Share on Google +

f Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

Located in a recently renovated area of town, this place is going to be around for a long time. It attracts a lot of

...

[Read more ▾](#)

**Brianna T.**

Rating: ★★★☆☆

SHARE THIS REVIEW

📧 Share on Google +

f Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

Man, I wish this place was better. I like the concept, and I like the people, but the food and drinks are really just sub-

...

[Read more ▾](#)

**Ella M.**

Rating: ★☆☆☆☆

SHARE THIS REVIEW

📧 Share on Google +

f Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

On the weekends, I usually check out different cafes and try to support local businesses. But this is one place I

...

[Read more ▾](#)

**Barry C.**

Rating: ★★★★★

SHARE THIS REVIEW

📧 Share on Google +

f Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

What a gem! Stumbled on this place when we were looking for a quick bite to eat and it was better than

...

[Read more ▾](#)

**Tina B.**

Rating: ★★★★★

SHARE THIS REVIEW

g+ Share on Google +

f Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

I think I found my new go-to place. Reasonably-priced tasty treats and awesome lattes. I was impressed by their

...

[Read more ▾](#)

**Victor H.**

Rating: ★☆☆☆☆

SHARE THIS REVIEW

g+ Share on Google +

f Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

Don't waste your money. The people were super disorganized and then they messed up my order. When I

...

[Read more ▾](#)

**Mike R.**

Rating: ★★☆☆☆

SHARE THIS REVIEW

g+ Share on Google +

f Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

Pretty disappointing. I was excited when it opened because the area could use some "local" businesses, but it

...

[Read more ▾](#)

**Priscilla P.**

Rating: ★★★★★☆

SHARE THIS REVIEW

g+ Share on Google +

f Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

Quality beans, friendly faces, and great prices! I started coming here after getting annoyed with some of the

...

[Read more ▾](#)

**Lyla A.**

Rating: ★★★★★☆

SHARE THIS REVIEW

g+ Share on Google +

f Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

As a café connoisseur, I can tell you that this place is extremely...average. Some things, like the drip coffee are

...

[Read more ▾](#)

## NOT RECOMMENDED

(We use automated software to recommend the reviews we think will be the most helpful to the Yelp community based primarily on quality, reliability and the reviewer's activity on Yelp. Advertisers get no special treatment. The reviews below didn't make the cut and are therefore not factored into this business's overall star rating. Check out our FAQ for more details. )

**Ariel W.**

**Rating:** ★★★★★

SHARE THIS REVIEW

 Share on Google +

 Share on Facebook

 Share on Twitter

 Email to a friend

You should definitely check out this place. Parking is good. Food is good. And the people are better. Its like their goal ...  
[Read more ▾](#)

Which review would you like to read in full?

- ☐ Rose S. (★★★★☆)
- ☐ Dan T. (★★★★☆)
- ☐ Brianna T. (★★☆☆☆)
- ☐ Ella M. (★☆☆☆☆)
- ☐ Barry C. (★★★★★)
- ☐ Tina B. (★★★★★)
- ☐ Victor H. (★☆☆☆☆)
- ☐ Mike R. (★★☆☆☆)
- ☐ Priscilla P. (★★★★☆)
- ☐ Lyla A. (★★★★☆)
- ☐ Ariel W. (★★★★★)

**Ella M.**

**Rating:** ★☆☆☆☆

SHARE THIS REVIEW

 Share on Google +

 Share on Facebook

 Share on Twitter

 Email to a friend

On the weekends, I usually check out different cafes and try to support local businesses. But this is one place I won't go back to. They weren't too friendly and the food was quite bad. I had a turkey wrap and wouldn't have known there was any turkey in it if not for opening it up.



**Victor H.**

Rating: ★☆☆☆☆

SHARE THIS REVIEW

📧 Share on Google +

📘 Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

Don't waste your money. The people were super disorganized and then they messed up my order. When I asked them to fix it, they acted as if it was my fault that they put stuff on my sandwich which I specifically requested be removed when I ordered.

**Brianna T.**

Rating: ★★☆☆☆

SHARE THIS REVIEW

📧 Share on Google +

📘 Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

Man, I wish this place was better. I like the concept, and I like the people, but the food and drinks are really just sub-par. I don't see myself going here again unless its someone else's choice. The town could really use a place like this if they can get their act together.

**Mike R.**

Rating: ★★☆☆☆

SHARE THIS REVIEW

📧 Share on Google +

📘 Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

Pretty disappointing. I was excited when it opened because the area could use some "local" businesses, but it really didn't differentiate itself well. I'm willing to pay a little more IF its better, but I don't think West End is good enough to justify me switching from other places.

**Lyla A.**

Rating: ★★★☆☆

SHARE THIS REVIEW

📧 Share on Google +

📘 Share on Facebook

🐦 Share on Twitter

✉ Email to a friend

As a café connoisseur, I can tell you that this place is extremely...average. Some things, like the drip coffee are done quite well, but others, like their lattes, are nothing to write home about. Its nice for a change of pace but not much else.

**Rose S.**

Rating: ★★☆☆☆

SHARE THIS REVIEW

- g+ Share on Google +
- f Share on Facebook
- Share on Twitter
- Email to a friend

I usually come here to get a sandwich when I'm in the area and hungry. I've tried a bunch of stuff on the menu but really only like a handful of things. If you want a sandwich, avoid their chicken (its pretty dry) but the turkey sandwiches are quite good.

**Dan T.**

Rating: ★★★★★

SHARE THIS REVIEW

- g+ Share on Google +
- f Share on Facebook
- Share on Twitter
- Email to a friend

Located in a recently renovated area of town, this place is going to be around for a long time. It attracts a lot of different people: students, business folks, and the like but they're able to have a menu that caters to everyone.

**Priscilla P.**

Rating: ★★★★★

SHARE THIS REVIEW

- g+ Share on Google +
- f Share on Facebook
- Share on Twitter
- Email to a friend

Quality beans, friendly faces, and great prices! I started coming here after getting annoyed with some of the national chains that have permeated the area, and I'm going to keep coming back. Support local!

**Barry C.**

Rating: ★★★★★

SHARE THIS REVIEW

- g+ Share on Google +
- f Share on Facebook
- Share on Twitter
- Email to a friend

What a gem! Stumbled on this place when we were looking for a quick bite to eat and it was better than advertised. Food was great, coffee was great, and the ambience was quite charming. I'll be back soon!

**Tina B.**

Rating: ★★★★★

SHARE THIS REVIEW

- Share on Google +
- Share on Facebook
- Share on Twitter
- Email to a friend

I think I found my new go-to place. Reasonably-priced tasty treats and awesome lattes. I was impressed by their selection of things and the employees were super friendly. Even though they were busy, they still asked about my day and made me feel special.

**Ariel W.**

Rating: ★★★★★

SHARE THIS REVIEW

- Share on Google +
- Share on Facebook
- Share on Twitter
- Email to a friend

You should definitely check out this place. Parking is good. Food is good. And the people are better. Its like their goal is to treat each customer like family. I was very satisfied with their level of service and quality of their product.

Which review would you like to read next?

- ☐ Rose S. (★★★★☆)
- ☐ Dan T. (★★★★☆)
- ☐ Brianna T. (★★★☆☆)
- ☐ Ella M. (★☆☆☆☆)
- ☐ Barry C. (★★★★★)
- ☐ Tina B. (★★★★★)
- ☐ Victor H. (★☆☆☆☆)
- ☐ Mike R. (★★★☆☆)
- ☐ Priscilla P. (★★★★☆)
- ☐ Lyla A. (★★★★☆)
- ☐ Ariel W. (★★★★★)
- ☐ I'M FINISHED READING REVIEWS



Now that you have finished reading the reviews, we have some questions about the cafe.

Based on the information you have seen, what star rating would you give this cafe? That is, how many stars do you think truly reflects the quality of West End Cafe?

My rating:

To what degree would you say that your rating of West End Cafe is accurate?

Not at all ☐ ☐ ☐ ☐ ☐ ☐ ☐ Extremely

Given the available information, how accurate do you think is the 3.5 out of 5.0 average star rating from other consumers for West End Cafe?

Not at all ☐ ☐ ☐ ☐ ☐ ☐ ☐ Extremely

---

Now we have a few additional questions about West End Cafe.

How strongly would you consider visiting this Cafe?

I would definitely not consider visiting here ☐ ☐ ☐ ☐ ☐ ☐ I would strongly consider visiting here

Not at all likely                      Extremely likely

O O O O O  
O O

Much Less ○ ○ ○ ○ ○ ○ ○ Much More

Extremely Low Risk      ○      ○      ○      ○      ○      Extremely High Risk

Extremely Low Quality      ○      ○      ○      ○      ○      Extremely High Quality

Now we'd like you to think about the reviews you read. Specifically, we'd like you to think about both the negative and positive reviews.

How likely is it that the reviews you read were trustworthy?

	Not at all						Extremely
Negative Reviews	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Positive Reviews	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How likely is it that the reviews you read were honest?

	Not at all						Extremely
Negative Reviews	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Positive Reviews	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How likely is it that the reviews you read were accurate?

	Not at all						Extremely
Negative Reviews	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Positive Reviews	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How much would you like to punish West End Cafe?

Not at all						Very Much
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How much would you like to reward West End Cafe?

Not at all						Very Much
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Have you ever encountered a message disclosing that fraudulent reviews have been identified before?

- ☐ No
- ☐ Yes

How familiar are you with each of the following:

	Not at all				Extremely
West End Cafe	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Yelp.com	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How often do you:

	Never				Often
visit West End Cafe	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
visit Yelp.com	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>


*Appendix C3 – Study 4 Stimuli*

## West End Hotel

★★★★☆ (264 reviews)

3.2 out of 5 stars

\$\$ - Hotels – [Hotel Website](#)



**Alert**



Yelp caught a competitor paying others to write negative reviews for this business. This means that someone was trying to artificially decrease the ratings of this business. We weren't fooled, but wanted you to know because these actions not only hurt consumers, but also honest businesses who play by the rules.

## West End Hotel

★★★★☆ (264 reviews)

3.2 out of 5 stars

\$\$ - Hotels - [Hotel Website](#)



### Alert



Yelp caught a competitor paying others to write negative reviews for this business. This means that someone was trying to artificially decrease the ratings of this business. We weren't fooled, but wanted you to know because these actions not only hurt consumers, but also honest businesses who play by the rules.

We have removed the fraudulent reviews, resulting in a 3.2 out of 5.0 star rating, based on 264 reviews.

## West End Hotel

★★★★☆ (264 reviews)

3.2 out of 5 stars

\$\$ - Hotels - [Hotel Website](#)



### Alert



Yelp caught this business paying others to write positive reviews for this business. This means that someone was trying to artificially increase the ratings of this business. We weren't fooled, but wanted you to know because these actions not only hurt consumers, but also honest businesses who play by the rules.

## West End Hotel



3.2 out of 5 stars

\$\$ - Hotels - [Hotel Website](#)



### Alert



Yelp caught this business paying others to write positive reviews for this business. This means that someone was trying to artificially increase the ratings of this business. We weren't fooled, but wanted you to know because these actions not only hurt consumers, but also honest businesses who play by the rules.

We have removed the fraudulent reviews, resulting in a 3.2 out of 5.0 star rating, based on 264 reviews.

*[reviews and measures follow Study 2 Round 1]*



## TABLES

Table 1 – Essay I Summary of Designs and Measures

Study	Product	Option	Average Product Ratings					Review Volumes					Reported Measures			
			(1.)	(2.)	(3.)	(4.)	(5.)	Low	Moderately Low	Moderately High	High	Absent	Relative Preference	Absolute Choice	Choice Deferral	Process Measures
1	Over-the-Ear Headphones	A	-	-	.5	-	-	8	72	201	456	Absent	x	-	-	-
		B	-	-	.3	-	-	64	128	257	512	Absent				
	Coffee Makers	A	-	-	.4	-	-	6	64	180	412	Absent	x	-	-	-
		B	-	-	.0	-	-	58	116	232	464	Absent				
	Microwaves	A	-	-	.5	-	-	9	71	195	443	Absent	x	-	-	-
		B	-	-	.2	-	-	62	124	248	496	Absent				
	Speaker Systems	A	-	-	.4	-	-	12	86	234	530	Absent	x	-	-	-
		B	-	-	.1	-	-	74	148	296	592	Absent				
	Lounge Chairs	A	-	-	.7	-	-	5	103	299	691	Absent	x	-	-	-
		B	-	-	.4	-	-	98	196	392	784	Absent				
2	Camping Lamps	A	-	-	.2	-	-	61	-	-	361	Absent	-	x	x	-
		B	-	-	.6	-	-	22	-	-	322	Absent				
		C	-	-	.8	-	-	5	-	-	305	Absent				
		D	-	-	.4	-	-	43	-	-	343	Absent				
3	Blenders	A	-	.4	.4	.4	-	8	-	-	408	-	X	-	-	-
		B	-	.1	.1	.1	-	64	-	-	464	-				
4	Earbud Headphones	A	-	-	.8, .6, .4, .2	.8, .6, .4, .2	-	9	-	-	409	-	x	-	-	-
		B	-	-	.6, .4, .2, .0	.6, .4, .2, .0	-	57	-	-	457	-				
5	Hand Mixers	A	.3 (.4)	-	-	(.9)	.0	13	-	-	313	-	x	-	-	-
		B	.0 (.1)	-	-	.7, (.6)	-	34	-	-	334	-				
6	Blenders	A	-	-	.4	-	-	8	-	-	408	-	-	x	x	x
		B	-	-	.1	-	-	64	-	-	464	-				
7	Microwaves	A	-	-	.5	-	-	9	-	-	443	-	x	-	-	x
		B	-	-	.2	-	-	62	-	-	496	-				

**Note:** For Study 4, all pairwise comparisons were used for average product ratings where A was greater than B, resulting in 10 different comparisons. For Study 5, numbers in parentheses were paired together in their respective negative and positive conditions.

Table 2 – Essay I Means Summary across Studies

	Average Product Rating		Review Volumes				
	Valence	Difference	Low	Moderately Low	Moderately High	High	Absent
Study 1 (n = 250)	Neutral	-	3.96	2.93	2.83	2.63	2.81
Study 2 (n = 144)	Neutral	-	49	-	-	78	76
Study 3 (n = 433)	Negative	-	4.15	-	-	3.32	-
	Neutral	-	4.38	-	-	3.38	-
	Positive	-	3.04	-	-	2.97	-
Study 4 (n = 410)	Negative	-	3.58	-	-	3.67	-
	Negative (extreme)	-	3.84	-	-	3.44	-
	Positive	-	4.08	-	-	2.71	-
	Positive (extreme)	-	3.48	-	-	2.96	-
Study 5 (n = 705)	Neutral	-	3.98	-	-	2.77	-
	Positive	-	3.32	-	-	2.66	-
	-	Small	3.62	-	-	2.98	-
	-	Large	3.68	-	-	2.45	-
Study 6 (n = 143)	Neutral	-	36	-	-	61	61
Study 7 (n = 92)	Neutral	-	4.09	-	-	3.68	-

Notes: Studies 1, 3, 4, 5, 7 use a relative preference measure (1 = Strongly prefer higher-rated, fewer reviews Option, 7 = Strongly prefer lower-rated, more reviews option). Studies 2 and 6 use an absolute choice measure where the number reported is the percentage choosing the higher-rated, fewer reviews choice option.

Table 3 – Essay I Study 7: Eye-tracking Attribute Transition Matrices by Review Volume Level

<b>LOW REVIEW VOLUMES</b>	<i>To Image</i>	<i>To Brand &amp; Price</i>	<i>To Rating</i>	<i>To Volume</i>	<i>To Add'l Info</i>	<i>To End</i>
<i>From Image</i>	68.23%	14.47%	4.51%	2.26%	6.02%	4.51%
<i>From Brand &amp; Price</i>	8.94%	72.81%	12.23%	2.37%	2.92%	0.73%
<i>From Rating</i>	3.50%	10.07%	62.58%	19.69%	3.72%	0.44%
<i>From Volume</i>	2.65%	3.41%	24.24%	49.62%	18.94%	1.14%
<i>From Add'l Info</i>	9.66%	1.87%	1.71%	2.80%	82.09%	1.87%

<b>HIGH REVIEW VOLUMES</b>	<i>To Image</i>	<i>To Brand &amp; Price</i>	<i>To Rating</i>	<i>To Volume</i>	<i>To Add'l Info</i>	<i>To End</i>
<i>From Image</i>	71.74%	12.89%	3.64%	2.15%	6.12%	3.47%
<i>From Brand &amp; Price</i>	9.35%	73.28%	12.02%	1.53%	2.48%	1.34%
<i>From Rating</i>	4.26%	9.31%	64.10%	14.63%	6.12%	1.60%
<i>From Volume</i>	4.69%	2.17%	13.36%	60.65%	17.33%	1.81%
<i>From Add'l Info</i>	8.20%	1.64%	1.37%	4.37%	83.47%	0.96%

Table 4 – Essay II Summary of Hypotheses Tested in Each Study

	<b>Hypotheses Tested</b>	<b>Results</b>
<b>Study 1</b>	H1 (partial)	Demonstrated the effect of a positive fake review alert on brand ratings
<b>Study 2</b>	H1 – H3, H5	Demonstrated the effect of positive and negative fake review alerts on brand ratings, average valence of reviews read, and the desire for justice
<b>Study 3</b>	H1 – H5	Replicated prior findings while demonstrating the mediating role of desire for justice in the effect of fake review alerts on brand ratings
<b>Study 4</b>	H1 – H6	Replicated prior findings while also demonstrating the moderating role of salient aggregate information

## FIGURES

Figure 1 – Essay I Conceptual Model

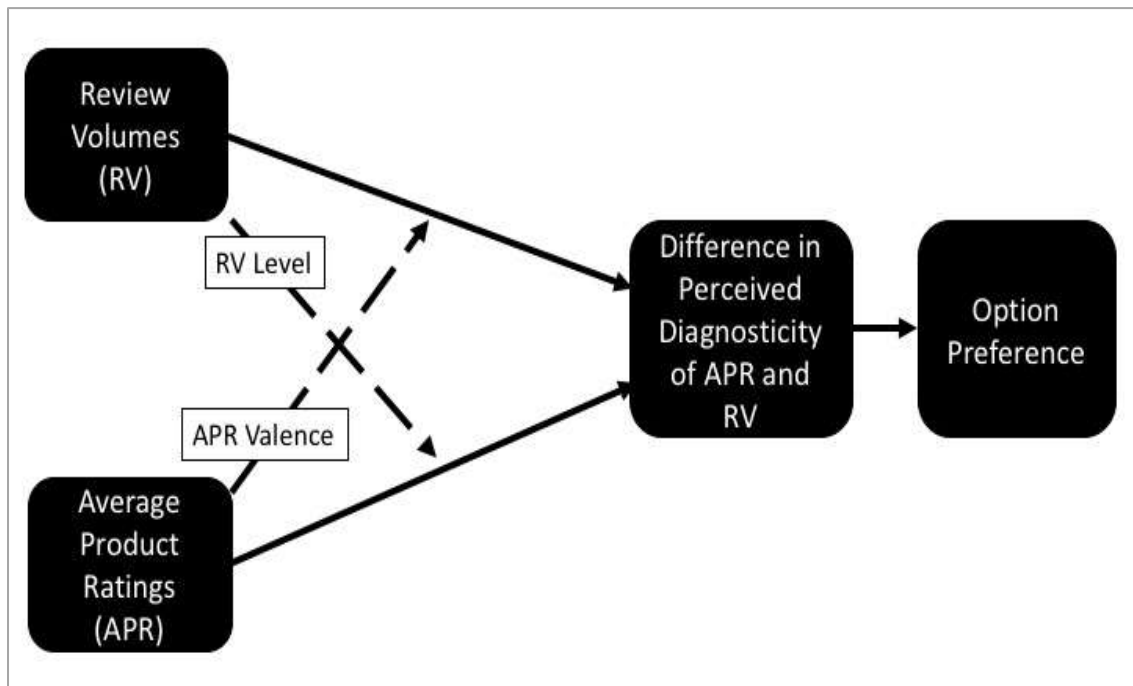


Figure 2 – Essay I Example of the Standard Study Stimuli

[SHOP ALL DEPARTMENTS -](#)

SEARCH:

⌂

CART (0)

[Home](#) > [Compare Products](#)

Compare Products

Remove [x]

A

Add to Cart

Name	Cuisinart® Rotisserie Convection Toaster Oven
Price	\$169.98
Rating	3.8 / 5.0
# of Reviews	408

Remove [x]

B

Add to Cart

Name	KitchenAid® Convection Toaster Oven
Price	\$168.99
Rating	3.0 / 5.0
# of Reviews	454

Figure 3 – Essay I Study 1 Preference Results

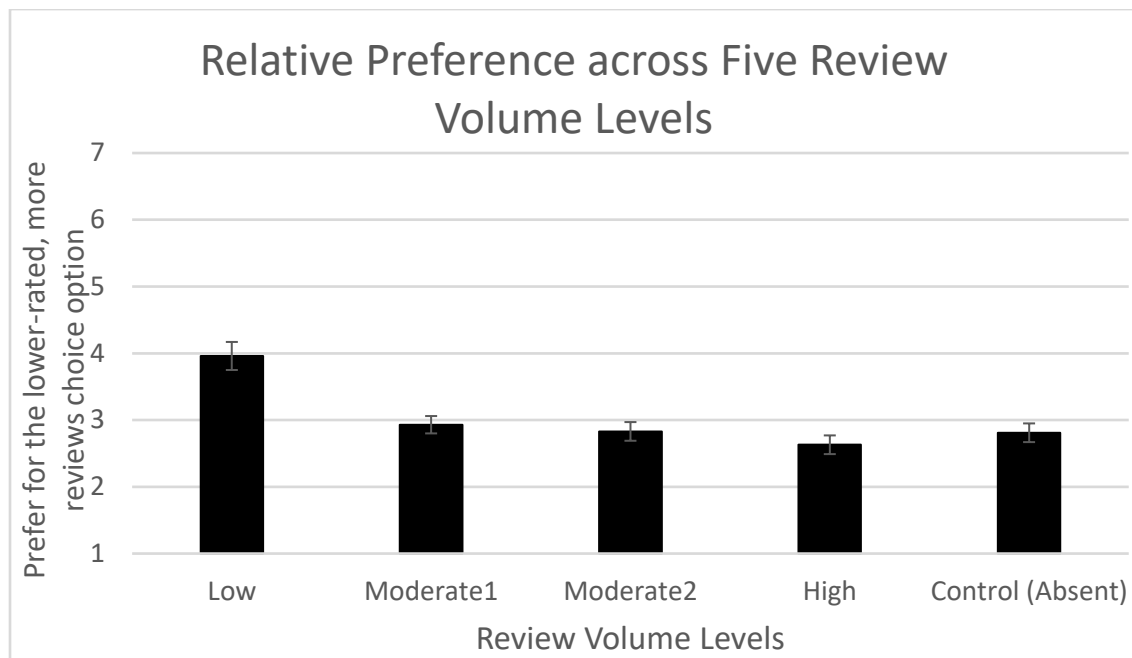


Figure 4 – Essay I Study 3 Preference Results

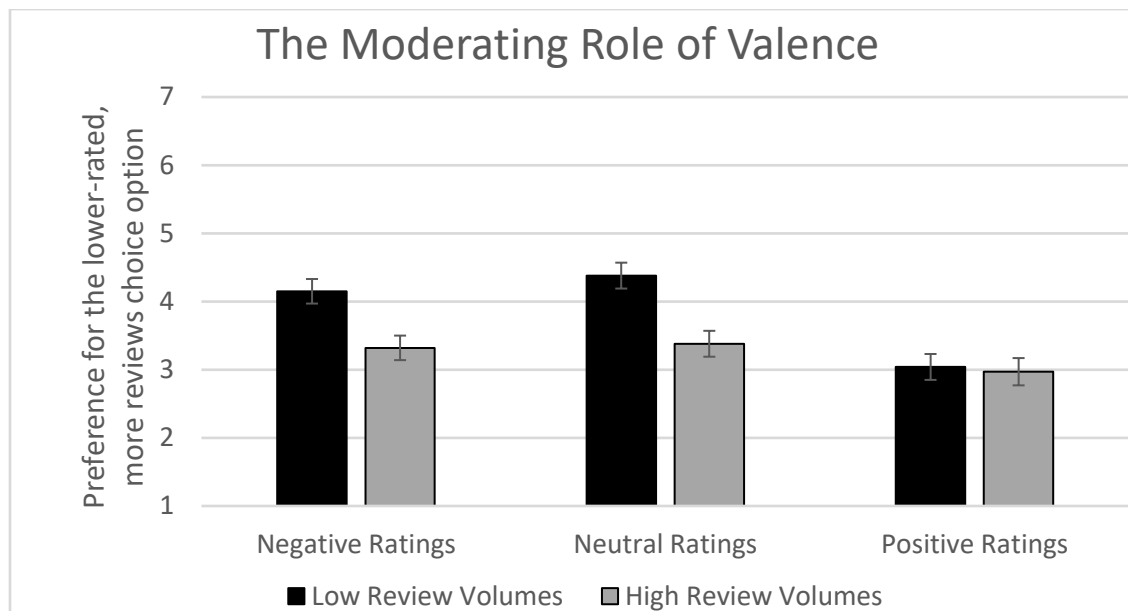




Figure 5 – Essay I Study 4 Preference Results: Volume by Valence Interaction

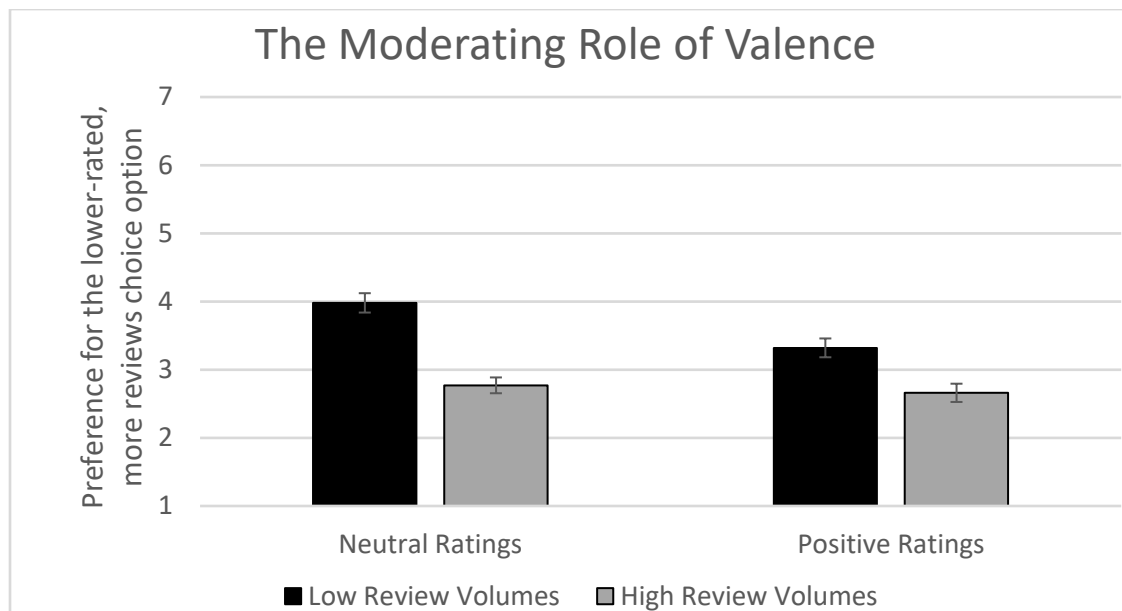


Figure 6 – Essay I Study 4 Preference Results: Volume by Difference Size Interaction

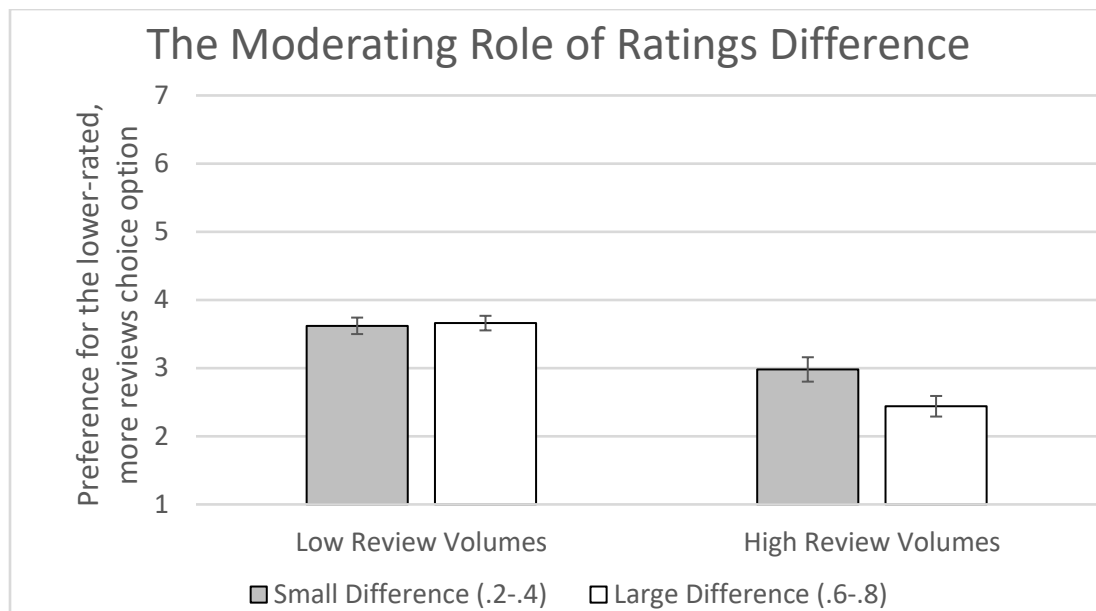
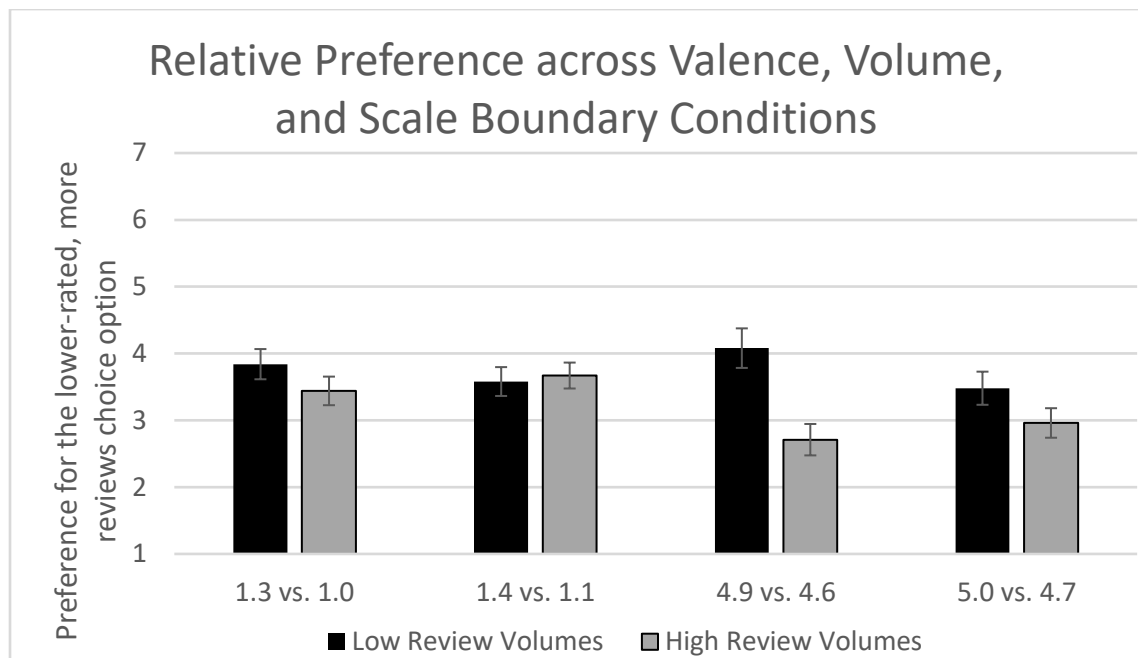


Figure 7 – Essay I Study 5 Preference Results



Figures 8A-D – Essay I Implications Simulations for Proactive and Reactive Strategies

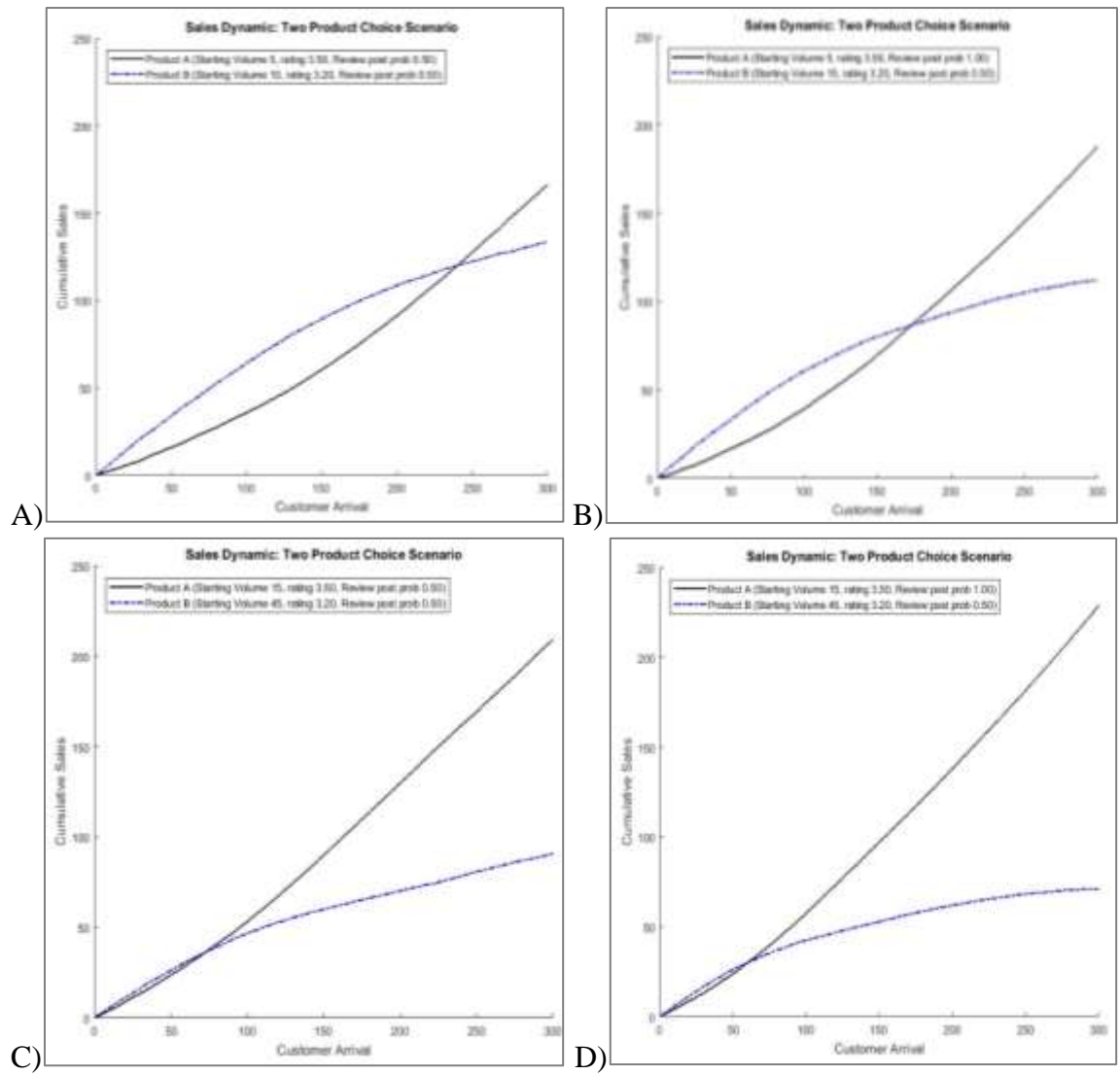


Figure 9 – Essay II Study 1: Average Brand Ratings Before, During, and After a Fake Positive Review Alert



Figure 10 – Essay II Study 2: Brand Ratings



Figure 11 – Essay II Study 2: Perception of Average Rating Accuracy

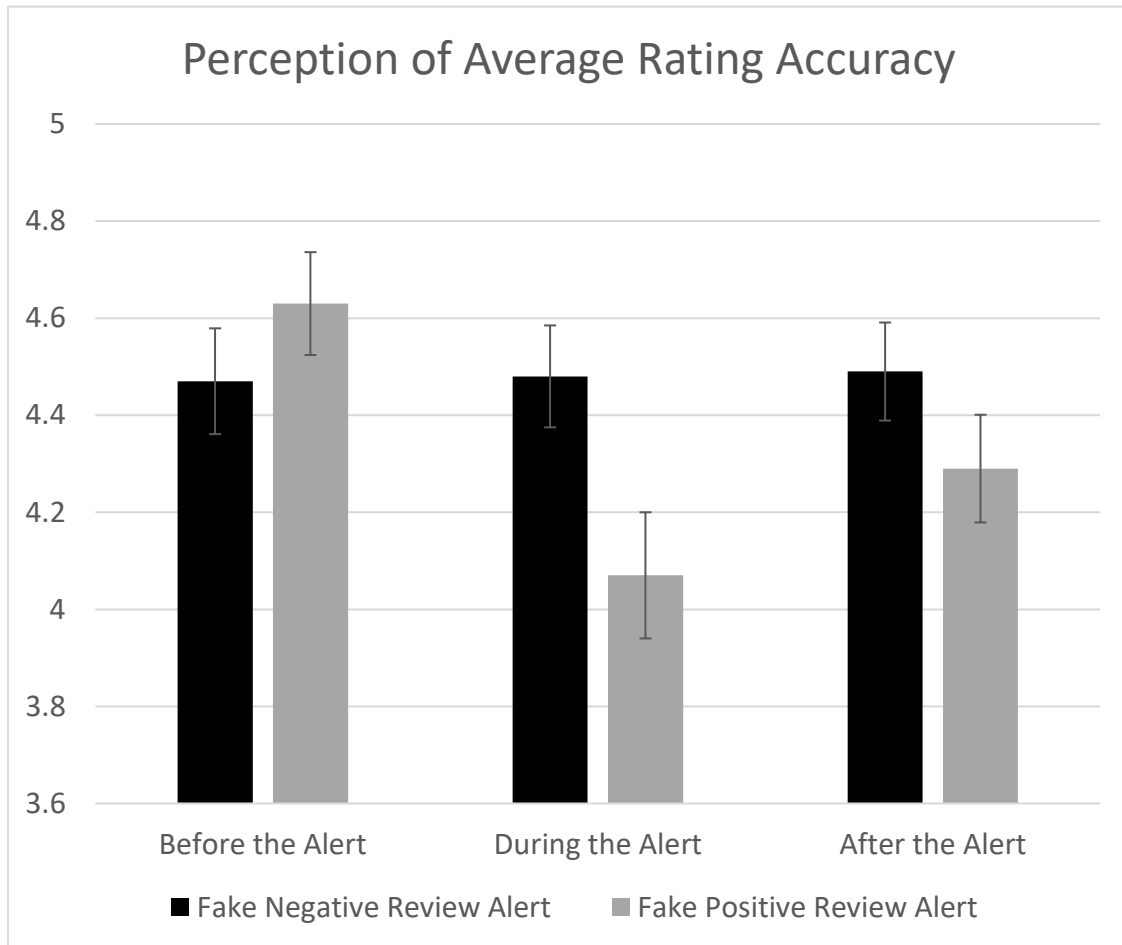


Figure 12 – Essay II Study 2: Desire for Justice





Figure 13 – Essay II Study 2: Average Valence of Reviews Read



Figure 14 – Essay II Study 3: Brand Ratings

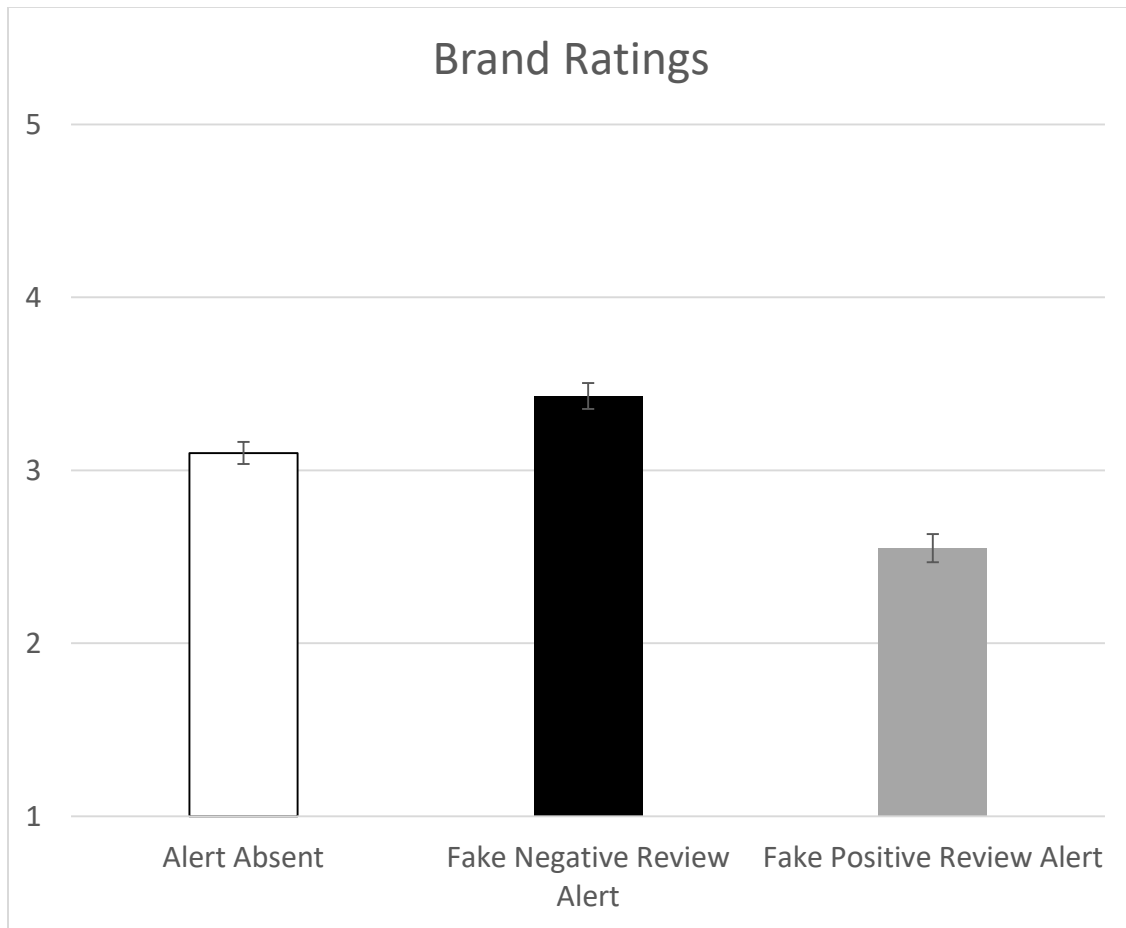


Figure 15 – Essay II Study 3: Perception of Average Rating Accuracy

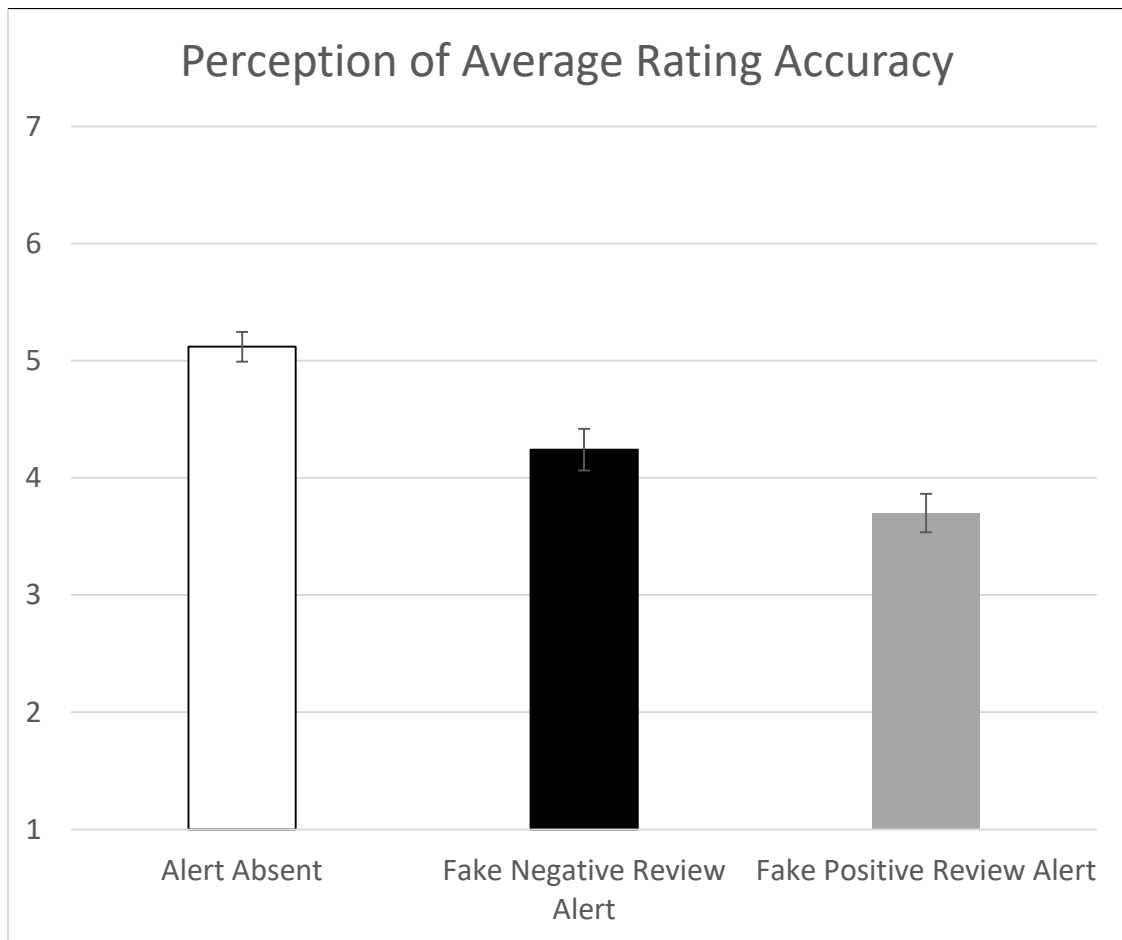


Figure 16 – Essay II Study 3: Desire for Justice



Figure 17 – Essay II Study 3: Average Valence of Reviews Read

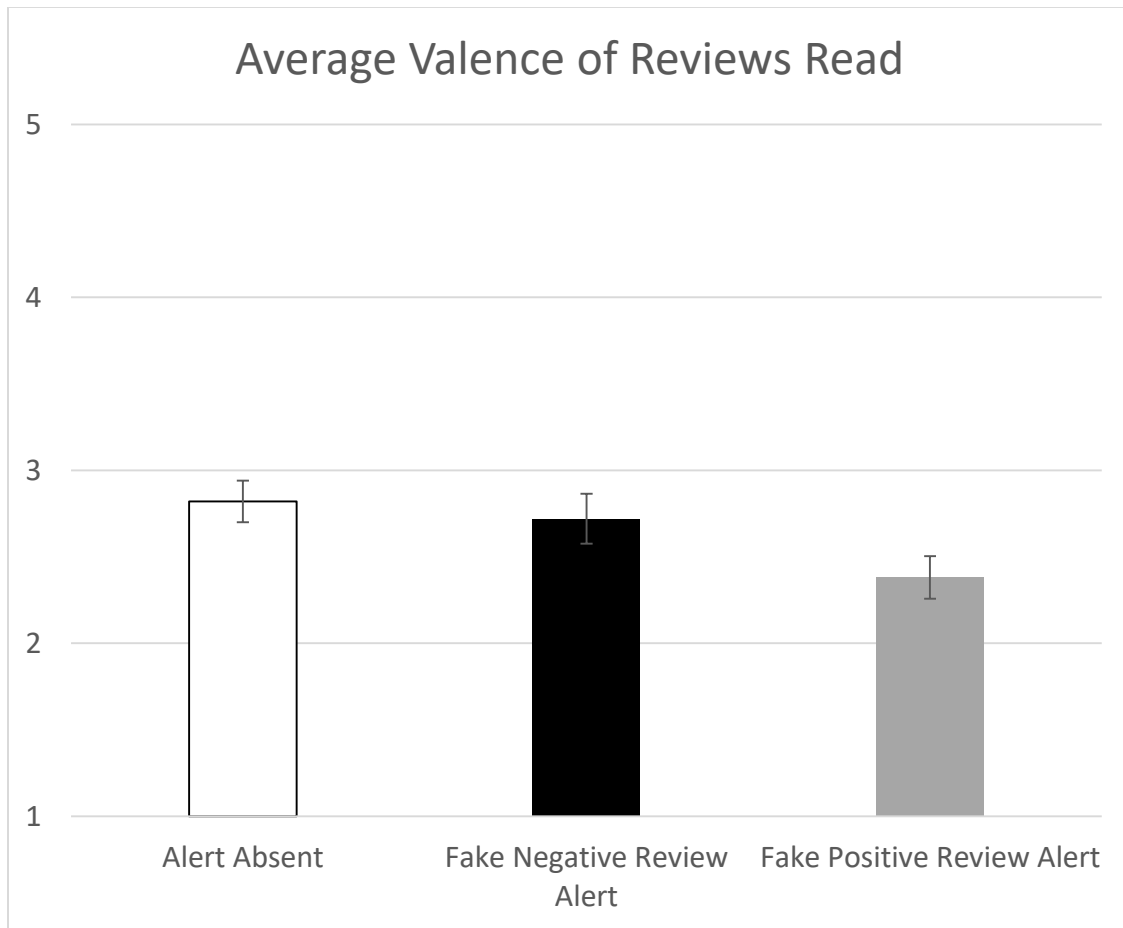


Figure 18 – Essay II Study 4: Brand Ratings

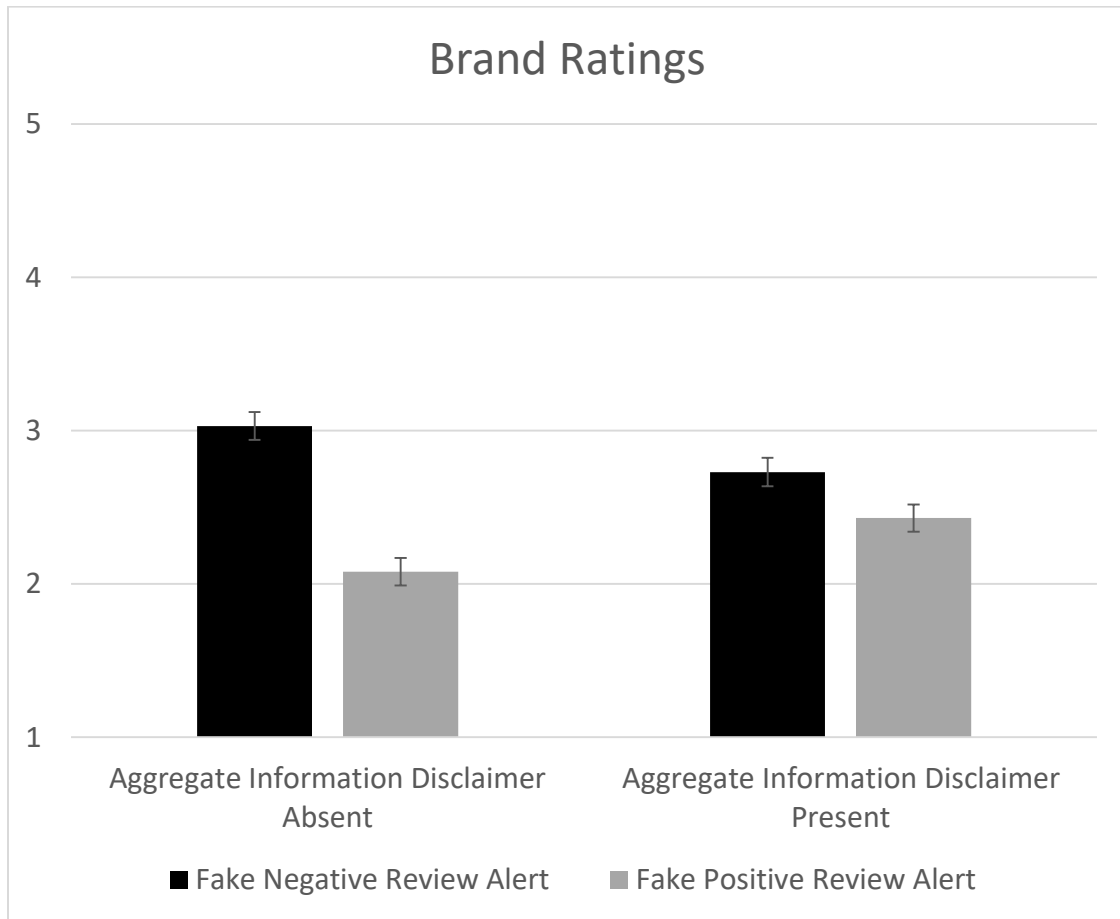


Figure 19 – Essay II Study 4: Perception of Average Rating Accuracy

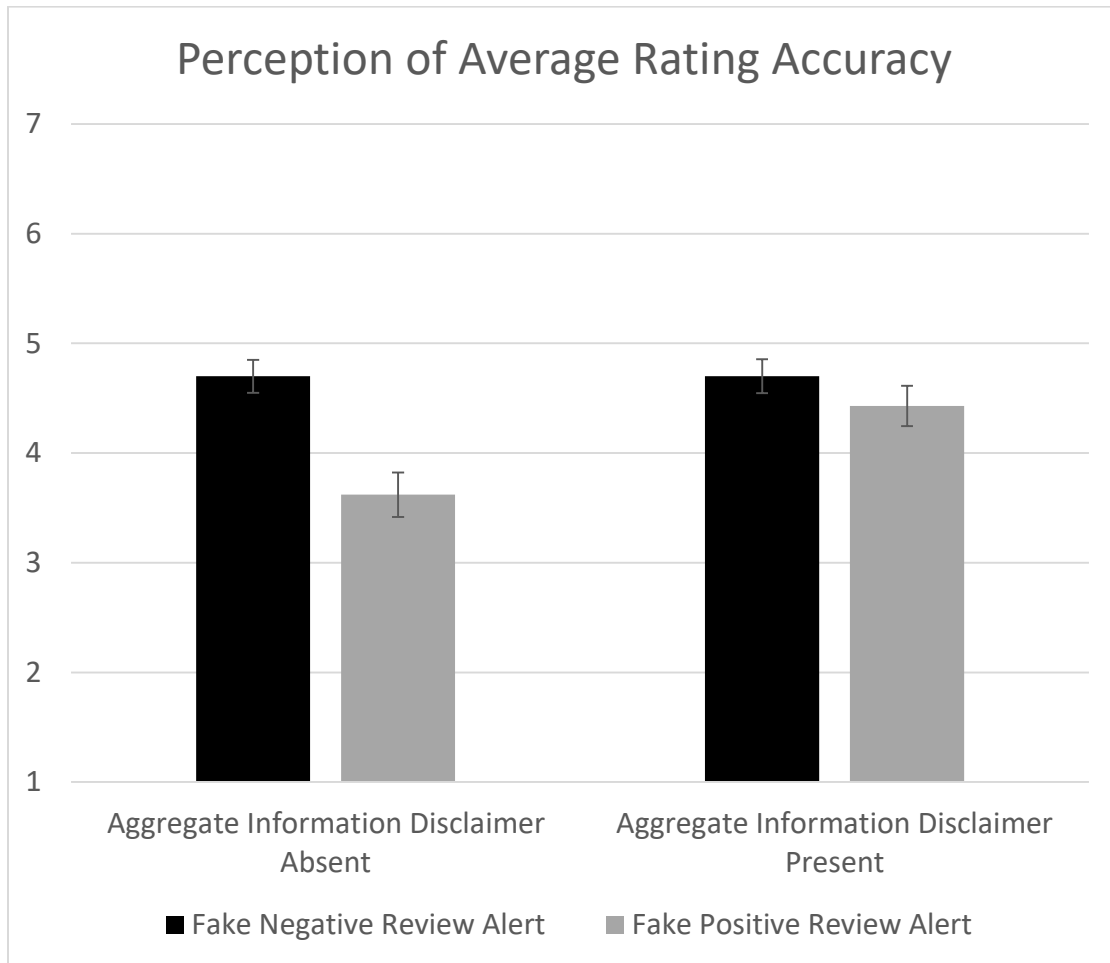


Figure 20 – Essay II Study 4: Desire for Justice





## REFERENCES

- Akoglu, Leman, Rishi Chandy, and Christos Faloutsos (2013), "Opinion Fraud Detection in Online Reviews by Network Effects," *ICWSM*, 13, 2-11.
- Bagchi, Rajesh, and Xingbo Li (2010), "Illusionary Progress in Loyalty Programs: Magnitudes, Reward Distances, and Step-Size Ambiguity," *Journal of Consumer Research*, 37 (August), 888-901.
- Banerjee, Abhijit V. (1992), "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics*, 107 (3), 797-817.
- Banks, Sandy (2013), "Putting Yelp in the Rear-Review Mirror," (accessed October 18, 2017), [available at <http://articles.latimes.com/2013/apr/22/local/la-me--banks-yelp-backlash-20130423>].
- Baumeister, Roy F., Ellen Bratslavsky, Catrin Finkenauer, and Kathleen D. Vohs (2001), "Bad is Stronger than Good," *Review of General Psychology*, 5 (December), 323.
- Bechwati, Nada Nasr, and Maureen Morrin (2003), "Outraged Consumers: Getting Even at the Expense of Getting a Good Deal," *Journal of Consumer Psychology*, 13(4), 440-453.
- Bies, Robert J., Thomas M. Tripp, and Roderick M. Kramer (1997), "At the Breaking Point: Cognitive and Social Dynamics of Revenge in Organizations."
- Bockenholt, Ulf, Dietrich Albert, Michael Aschenbrenner, and Franz Schmalhofer (1991), "The Effects of Attractiveness, Dominance, and Attribute Differences on Information Acquisition in Multiattribute Binary Choice," *Organizational Behavior and Human Decision Processes*, 49 (August), 258-281.
- Bonifield, Carolyn, and Catherine Cole (2007) "Affective Responses to Service Failure: Anger, Regret, and Retaliatory versus Conciliatory Responses," *Marketing Letters*, 18(1-2), 85-99.
- BrightLocal (2017), "Local Consumer Review Survey," (accessed February 24, 2018), [available at <https://www.brightlocal.com/learn/local-consumer-review-survey/>].
- Burson, Katherine A., Richard P. Larrick, and John G. Lynch Jr. (2009), "Six of One, Half Dozen of the Other: Expanding and Contracting Numerical Dimensions Produces Preference Reversals," *Psychological Science*, 20 (September), 1074-1078.
- Campbell, Margaret C. (1995), "When Attention-Getting Advertising Tactics Elicit Consumer Inferences of Manipulative Intent: The Importance of Balancing Benefits and Investments," *Journal of Consumer Psychology*, 4(3), 225-254.
- , and Amna Kirmani (2000), "Consumers' Use of Persuasion Knowledge: The Effects of Accessibility and Cognitive Capacity on Perceptions of an Influence Agent," *Journal of Consumer Research*, 27(1), 69-83.
- Chandon, Pierre, and Nailya Ordabayeva (2017), "The Accuracy of Less: Natural Bounds Explain Why Quantity Decreases are Estimated More Accurately than Quantity Increases," *Journal of Experimental Psychology: General* 146 (February), 250.
- Chatterjee, Subimal, and Timothy B. Heath (1996), "Conflict and Loss Aversion in Multiattribute Choice: The Effects of Trade-off Size and Reference Dependence on Decision Difficulty," *Organizational Behavior and Human Decision Processes*, 67 (August), 144-155.
- Chen, Pei-Yu, Samita Dhanasobhon, and Michael D. Smith (2008), "All Reviews are Not Created Equal: The Disaggregate Impact of Reviews and Reviewers at Amazon.com".

- Chen, Yu-Jen, and Amna Kirmani (2015), "Posting Strategically: The Consumer as an Online Media Planner," *Journal of Consumer Psychology*, 25(4), 609-621.
- Chevalier, Judith A., and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (August), 345-354.
- Chintagunta, Pradeep K., Shyam Gopinath, and Sriram Venkataraman (2010), "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets," *Marketing Science*, 29 (September), 944-957.
- Clemons, Eric K., Guodong Gordon Gao, and Lorin M. Hitt (2006), "When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry," *Journal of Management Information Systems*, 23 (October), 149-171.
- Curtis, Sophie (2014), "Yelp CEO: How 'Sting' Operations are Cleaning up Reviews on the Web," (accessed October 18, 2017), [available at <http://www.telegraph.co.uk/technology/news/10571494/Yelp-CEO-how-sting-operations-are-cleaning-up-reviews-on-the-web.html>].
- De Langhe, Bart, Philip M. Fernbach, and Donald R. Lichtenstein (2015), "Navigating by the Stars: Investigating the Actual and Perceived Validity of Online User Ratings," *Journal of Consumer Research*, 42 (September), 817-833.
- Dellarocas, Chrysanthos, Xiaoquan Michael Zhang, and Neveen F. Awad (2007), "Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures," *Journal of Interactive Marketing*, 21(4), 23-45.
- Dhar, Ravi, and Steven J. Sherman (1996), "The Effect of Common and Unique Features in Consumer Choice," *Journal of Consumer Research*, 23 (December), 193-203.
- , and Stephen M. Nowlis (1999), "The Effect of Time Pressure on Consumer Choice Deferral," *Journal of Consumer Research*, 25 (March), 369-384.
- , and Itamar Simonson (2003), "The Effect of Forced Choice on Choice," *Journal of Marketing Research*, 40 (May), 146-160.
- Duan, Wenjing, Bin Gu, and Andrew B. Whinston (2008), "Do Online Reviews Matter?—An Empirical Investigation of Panel Data," *Decision Support Systems*, 45(4), 1007-1016.
- Etkin, Jordan, and Anastasiya Pocheptsova Ghosh (2017), "When Being in a Positive Mood Increases Choice Deferral," *Journal of Consumer Research*, 11 (December).
- Feldman, Jack M., and John G. Lynch (1988), "Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior," *Journal of Applied Psychology*, 73 (August), 421-435.
- Feng, Song, Ritwik Banerjee, and Yejin Choi (2012), "Syntactic Stylometry for Deception Detection," *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*, Association for Computational Linguistics.
- Fisher, Matthew, George E. Newman, Ravi Dhar, and Stijn van Osselaer (2018), "Seeing Stars: How the Binary Bias Distorts the Interpretation of Customer Ratings," *Journal of Consumer Research*, (forthcoming).
- Floyd, Kristopher, Ryan Freling, Saad Alhoqail, Hyun Young Cho, and Traci Freling (2014), "How Online Product Reviews Affect Retail Sales: A Meta-analysis," *Journal of Retailing*, 90(2), 217-232.

- Friestad, Marian, and Peter Wright (1994), "The Persuasion Knowledge Model: How People Cope with Persuasion Attempts," *Journal of Consumer Research*, 21(1), 1-31.
- Gilbert, Daniel T (1991), "How Mental Systems Believe," *American Psychologist*, 46 (February), 107.
- Godes, David, and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23 (January), 545-560.
- González-Vallejo, Claudia, and Elizabeth Moran (2001), "The Evaluability Hypothesis Revisited: Joint and Separate Evaluation Preference Reversal as a Function of Attribute Importance," *Organizational Behavior and Human Decision Processes*, 86 (November), 216-233.
- Grégoire, Yany, Thomas M. Tripp, and Renaud Legoux (2009), "When Customer Love Turns into Lasting Hate: The Effects of Relationship Strength and Time on Customer Revenge and Avoidance," *Journal of Marketing*, 73(6), 18-32.
- Herr, Paul M., Frank R. Kardes, and John Kim (1991), "Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective," *Journal of Consumer Research*, 17 (March), 454-462.
- Ho-Dac, Nga N., Stephen J. Carson, and William L. Moore (2013), "The Effects of Positive and Negative Online Customer Reviews: Do Brand Strength and Category Maturity Matter?," *Journal of Marketing*, 77(6), 37-53.
- Hsee, Christopher K. (1996), "The Evaluability Hypothesis: An Explanation for Preference Reversals Between Joint and Separate Evaluations of Alternatives," *Organizational Behavior and Human Decision Processes*, 67 (September), 247-257.
- , George F. Loewenstein, Sally Blount, and Max H. Bazerman (1999), "Preference Reversals between Joint and Separate Evaluations of Options: A Review and Theoretical Analysis," *Psychological Bulletin*, 125 (September), 576.
- (2000), "Attribute Evaluability and its Implications for Joint-Separate Evaluation Reversals and Beyond."
- , and Jiao Zhang (2010), "General Evaluability Theory," *Perspectives on Psychological Science*, 5 (July), 343-355.
- Isaac, Mathew S., and Kent Grayson (2017), "Beyond Skepticism: Can Accessing Persuasion Knowledge Bolster Credibility?," *Journal of Consumer Research*, 43(6), 895-912.
- Ito, Tiffany A., Jeff T. Larsen, N. Kyle Smith, and John T. Cacioppo (1998), "Negative Information Weighs More Heavily on the Brain: The Negativity Bias in Evaluative Categorizations," *Journal of Personality and Social Psychology*, 75 (October), 887.
- Jain, Shailendra Pratap, and Steven S. Posavac (2004), "Valenced Comparisons," *Journal of Marketing Research*, 41(1), 46-58.
- Johnson, Eric J., and Robert J. Meyer (1984), "Compensatory Choice Models of Noncompensatory Processes: The Effect of Varying Context," *Journal of Consumer Research*, 11 (June), 528-541.
- Kahneman, Daniel, and Amos Tversky (1979), "Prospect Theory: An Analysis of Decision under Risk," *Econometrica: Journal of the Econometric Society*, 263-291.
- Khare, Adwait, Lauren I. Labrecque, and Anthony K. Asare (2011), "The Assimilative and Contrastive Effects of Word-of-Mouth Volume: An Experimental Examination of Online Consumer Ratings," *Journal of Retailing*, 87 (March), 111-126.

- King, Dan, and Chris Janiszewski (2011), "The Sources and Consequences of the Fluent Processing of Numbers," *Journal of Marketing Research*, 48 (April), 327-341.
- Kirmani, Amna, and Akshay R. Rao (2000), "No Pain, No Gain: A Critical Review of the Literature on Signaling Unobservable Product Quality," *Journal of Marketing*, 64 (April), 66-79.
- , and Margaret C. Campbell (2004), "Goal Seeker and Persuasion Sentry: How Consumer Targets Respond to Interpersonal Marketing Persuasion," *Journal of Consumer Research*, 31(3), 573-582.
- , and Rui Zhu (2007), "Vigilant against Manipulation: The Effect of Regulatory Focus on the Use of Persuasion Knowledge," *Journal of Marketing Research*, 44(4), 688-701.
- Lembregts, Christophe, and Mario Pandelaere (2013), "Falling Back on Numbers: Quantitative Specifications as Sources of Control," *2013 Society for Consumer Psychology Annual Winter Conference*, (February).
- Liu, Yong (2006), "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing*, 70 (July), 74-89.
- Luca, Michael, and Georgios Zervas (2016), "Fake it Till you Make it: Reputation, Competition, and Yelp Review Fraud," *Management Science*, 62 (January), 3412-3427.
- Ludwig, Stephan, Ko De Ruyter, Mike Friedman, Elisabeth C. Brüggén, Martin Wetzels, and Gerard Pfann (2013), "More than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates," *Journal of Marketing*, 77 (January), 87-103.
- Lynch, John G., Howard Marmorstein, and Michael F. Weigold (1988), "Choices From Sets Including Remembered Brands: Use of Recalled Attributes and Prior Overall Evaluations," *Journal of Consumer Research*, 15 (September), 169-184.
- Mayzlin, Dina, Yaniv Dover, and Judith Chevalier (2014), "Promotional Reviews: An Empirical Investigation of Online Review Manipulation," *American Economic Review*, 104 (August), 2421-2455.
- McAuley, Julian, Rahul Pandey, and Jure Leskovec (2015), "Inferring Networks of Substitutable and Complementary Products," *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM (August), 785-794.
- McCullough, Michael E., Frank D. Fincham, and Jo-Ann Tsang (2003), "Forgiveness, Forbearance, and Time: The Temporal Unfolding of Transgression-related Interpersonal Motivations," *Journal of Personality and Social Psychology*, 84(3), 540.
- Moe, Wendy W., and Michael Trusov (2011), "The Value of Social Dynamics in Online Product Ratings Forums," *Journal of Marketing Research*, 48 (June), 444-456.
- Monga, Ashwani, and Rajesh Bagchi (2011), "Years, Months, and Days versus 1, 12, and 365: the Influence of Units versus Numbers," *Journal of Consumer Research*, 39 (September), 185-198.
- Mukherjee, Arjun, Vivek Venkataraman, Bing Liu, and Natalie S. Glance (2013), "What Yelp Fake Review Filter Might be Doing?," *ICWSM*.
- Murray, Keith B. (1991), "A Test of Services Marketing Theory: Consumer Information Acquisition Activities," *Journal of Marketing*, 10-25.

- National Retail Federation (2010), "E-Commerce Sales," (accessed November 28, 2016), [available at <http://research.nrffoundation.com>].
- Novemsky, Nathan, Ravi Dhar, Norbert Schwarz, and Itamar Simonson (2007), "Preference Fluency in Choice," *Journal of Marketing Research*, 44 (August), 347-356.
- Ott, Myle, Claire Cardie, and Jeffrey T. Hancock (2013), "Negative Deceptive Opinion Spam," *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.
- Pandelaere, Mario, Barbara Briers, and Christophe Lembregts (2011), "How to Make a 29% Increase Look Bigger: The Unit Effect in Option Comparisons," *Journal of Consumer Research*, 38 (February), 308-322.
- Payne, John W. (1976), "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis," *Organizational Behavior and Human Performance* 16 (August), 366-387.
- Petty, Richard E., and Duane T. Wegener (1993), "Flexible Correction Processes in Social Judgment: Correcting for Context-induced Contrast," *Journal of Experimental Social Psychology* 29(2), 137-165.
- Preacher, Kristopher J., Derek D. Rucker, and Andrew F. Hayes (2007), "Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions," *Multivariate Behavioral Research*, 42 (June), 185-227.
- Purohit, Devavrat, and Joydeep Srivastava (2001), "Effect of Manufacturer Reputation, Retailer Reputation, and Product Warranty on Consumer Judgments of Product Quality: A Cue Diagnosticity Framework," *Journal of Consumer Psychology*, 10 (January), 123-134.
- Rao, Akshay R., and Kent B. Monroe (1988), "The Moderating Effect of Prior Knowledge on Cue Utilization in Product Evaluations," *Journal of Consumer Research*, 15 (September), 253-264.
- , and ——— (1989), "The Effect of Price, Brand Name, and Store Name on Buyers' Perceptions of Product Quality: An Integrative Review," *Journal of Marketing Research*, (August), 351-357.
- Richardson, Paul S., Alan S. Dick, and Arun K. Jain (1994), "Extrinsic and Intrinsic Cue Effects on Perceptions of Store Brand Quality," *The Journal of Marketing*, (October), 28-36.
- Rozenkrants, Bella, S. Christian Wheeler, and Baba Shiv (2017), "Self-Expression Cues in Product Rating Distributions: When People Prefer Polarizing Products," *Journal of Consumer Research*, 44(4), 759-777.
- Rozin, Paul, and Edward B. Royzman (2001), "Negativity Bias, Negativity Dominance, and Contagion," *Personality and Social Psychology Review*, 5(4), 296-320.
- Schley, Dan R., Christophe Lembregts, and Ellen Peters (2017), "The Role of Evaluation Mode on the Unit Effect," *Journal of Consumer Psychology*, 27 (April), 278-286.
- Schlosser, Ann E. (2005), "Posting versus Lurking: Communicating in a Multiple Audience Context," *Journal of Consumer Research*, 32(2), 260-265.
- (2011), "Can Including Pros and Cons Increase the Helpfulness and Persuasiveness of Online Reviews? The Interactive Effects of Ratings and Arguments," *Journal of Consumer Psychology* 21(3), 226-239.

- Segal, David (2011), "A Rave, a Pan, or Just a Fake?" (accessed March 3, 2018), [available at <http://www.nytimes.com/2011/05/22/your-money/22haggler.html>].
- Slovic, Paul (1966), "Cue-Consistency and Cue-Utilization in Judgment," *The American Journal of Psychology*, 79 (September), 427-434.
- , and Sarah Lichtenstein (1971), "Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment," *Organizational Behavior and Human Performance*, 6 (November), 649-744.
- Sun, Monic (2012), "How Does the Variance of Product Ratings Matter?" *Management Science*, 58 (April), 696-707.
- Tax, Stephen S., Stephen W. Brown, and Murali Chandrashekar (1998), "Customer Evaluations of Service Complaint Experiences: Implications for Relationship Marketing," *Journal of Marketing*, 60-76.
- Thaler, Richard (1980), "Toward a Positive Theory of Consumer Choice," *Journal of Economic Behavior & Organization*, 1 (March), 39-60.
- Thompson, Suzanne C. (1981), "Will it Hurt Less if I Can Control it? A Complex Answer to a Simple Question," *Psychological Bulletin*, 90(1), 89.
- TripAdvisor (2018), "What will happen if a Business is found to Have Fraudulent Reviews?" (accessed February 5, 2018), [available at <https://www.tripadvisor.com/hc/en-gb/articles/200614957>].
- Tsiros, Michael, and Vikas Mittal (2000), "Regret: A Model of its Antecedents and Consequences in Consumer Decision Making," *Journal of Consumer Research*, 26 (March), 401-417.
- Tuttle, Brad (2015), "Amazon Lawsuit Shows that Fake Online Reviews are a Big Problem," (accessed March 3, 2018), [available at <http://time.com/money/4078632/amazon-fake-online-reviews/>].
- Tversky, Amos, and Daniel Kahneman (1985), "The Framing of Decisions and the Psychology of Choice," *Environmental Impact Assessment, Technology Assessment, and Risk Analysis*, Springer, Berlin, Heidelberg, 107-129.
- , and ——— (1995), "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5 (October), 297-323.
- , and Eldar Shafir (1992), "Choice under Conflict: The Dynamics of Deferred Decision," *Psychological Science* 3 (November), 358-361.
- Udo, Godwin J., and Gerald P. Marquis (2002), "Factors Affecting e-Commerce Web Site Effectiveness," *Journal of Computer Information Systems*, 42(2), 10-16.
- Villarroel Ordenes, Francisco, Stephan Ludwig, Ko De Ruyter, Dhruv Grewal, and Martin Wetzels (2017), "Unveiling what is written in the Stars: Analyzing Explicit, Implicit, and Discourse Patterns of Sentiment in Social Media," *Journal of Consumer Research*, 43 (April), 875-894.
- Ward, James C., and Amy L. Ostrom (2006), "Complaining to the Masses: The Role of Protest Framing in Customer-created Complaint Web Sites," *Journal of Consumer Research*, 33(2), 220-230.
- Wegener, Duane T., and Richard E. Petty (1995), "Flexible Correction Processes in Social Judgment: The Role of Naive Theories in Corrections for Perceived Bias," *Journal of Personality and Social Psychology*, 68(1), 36.

- , and ——— (1997), "The Flexible Correction Model: The Role of Naive Theories of Bias in Bias Correction," *Advances in Experimental Social Psychology*, 29, Academic Press, 141-208.
- Wilson, Timothy D., and Nancy Brekke (1994), "Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations," *Psychological Bulletin*, 116(1), 117.
- Woolacott, Emma (2017), "Amazon's Fake Review Problem Is Now Worse Than Ever, Study Suggests," (accessed March 3, 2018), [available at <https://www.forbes.com/sites/emmawoolacott/2017/09/09/exclusive-amazons-fake-review-problem-is-now-worse-than-ever/>].
- Yelp (2013), "How Yelp Protects Consumers from Fake Reviews," (accessed October 18, 2017), [available at <https://www.yelpblog.com/2013/05/how-yelp-protects-consumers-from-fake-reviews/>].
- You, Ya, Gautham G. Vadakkepatt, and Amit M. Joshi (2015) "A Meta-analysis of Electronic Word-of-Mouth Elasticity," *Journal of Marketing*, 79(2), 19-39.
- Zeelenberg, Marcel, and Rik Pieters (2007), "A Theory of Regret Regulation 1.0," *Journal of Consumer Psychology*, 17 (January), 3-18.
- Zhu, Feng, and Xiaoquan Zhang (2010), "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," *Journal of Marketing*, 74 (March), 133-148.